Guidelines for Quality Assurance and Quality Control in Surface Water Quality Programs in Alberta

Alberta Environment

Guidelines for Quality Assurance and Quality Control in Surface Water Quality Programs in Alberta

Prepared by:

Patricia Mitchell, M.Sc., P.Biol.

Patricia Mitchell Environmental Consulting

for

Alberta Environment

July 2006

W0603

Any comments, questions or suggestions regarding the content of this document may be directed to:

Environmental Monitoring and Evaluation Branch
Environmental Assurance Division
Alberta Environment
12 th Floor, Oxbridge Place
9820 – 106 Street
Edmonton, Alberta T5K 2J6
Fax: (780) 422-8606

Additional copies of this document may be obtained by contacting:

Information Centre
Alberta Environment
Main Floor, Oxbridge Place
9820 – 106 Street
Edmonton, Alberta T5K 2J6
Phone: (780) 427-2700
Fax: (780) 422-4086
Email: env.infocent@gov.ab.ca

PREFACE

Over the years, quality assurance for the acquisition of water quality data has become increasingly important. Development in Alberta is increasing at a rapid pace, but at the same time the people of Alberta are demanding that developers and the government protect the environment. Predevelopment studies and monitoring and assessment of impacts during and after development must be based on sound science. Sound science demands a good quality assurance program, so that government staff can demonstrate that the data collected are accurate and precise, ensuring that environmental decisions are valid.

The following document responds to the need to standardize quality assurance programs for all surface water quality studies conducted by and for the Alberta government. Although the document focuses on surface water and sediments, the principles would be similar for monitoring of groundwater, drinking water and other environmental studies. Very few similar documents are available in Canada. In the United States, the U.S. Environmental Protection Agency has extensive and detailed documents on quality assurance, mainly because any studies using federal funding are legally bound to follow quality assurance procedures.

Much of this guideline document is based on several literature sources. It should be remembered that in the field of quality assurance, very little is set in stone, and quite often one literature source contradicts another. Therefore, Alberta Environment decided to draft its own QA/QC guidelines, which professional staff agreed upon. This document should be reviewed and updated periodically as new information becomes available.

It is the intention of Alberta Environment to establish measurement quality objectives for variables of concern, to be based on recent data for all sampling programs. These will be published in a separate document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICIES

ACKNOWLEDGEMENTS

1. INTRODUCTION

Alberta Environment has sampled water and sediment quality in rivers, lakes and streams since the 1960s. The data collected are used to assess present conditions, compare data with water quality guidelines, investigate specific water quality issues, or determine long-term trends. Monitoring programs also assess whether regulatory processes are effective in protecting water bodies from excessive nutrients, metals, pesticides and toxic substances.

Another purpose in monitoring surface waters is in support of the *Water for Life* strategy of the Alberta government. *Water for Life* is a commitment to using scientific knowledge in decision-making to sustain water supplies, protect drinking water and ensure healthy aquatic ecosystems. Partnerships are a key direction in this initiative. The Surface Water Monitoring Subcommittee of the Water Quality Task Group has identified several key principles to ensure that good scientific information, which cannot be compromised, is collected. These include the following steps (Anderson et al. 2005):

Step 1: Scoping and Design
Competent program design requires a clear scientific understanding of the issues, the study objectives, appropriate methods and the natural dynamics of rivers, streams, lakes, wetlands and reservoirs.

Step 2: Sample Collection
Sample collection requires expertise and skill, including adherence to well-defined methods, good data management standards and health and safety considerations.

Step 3: Sample Analysis
Chemical, biological and physical analyses must be performed by competent laboratories and results must meet scientific criteria for the acceptability of results.

Step 4: Data Validation
High quality, reliable data must be ensured before they are stored electronically. This is done by confirming field and lab methods, checking results and QC data and ensuring proper coding.

Step 5: Data Storage
Information and data must be reliably stored over the long-term, and be easily accessible to all parties. It is undesirable to have many independent databases with separate validation procedures.

Step 6: Reporting
Competent parties should convert the data to accurate, reliable and scientifically defensible information in a timely manner.

All of the sampling programs conducted by Alberta Environment (AENV) depend on reliable and accurate data. The consequences of using poor quality data include faulty decisions, higher risk to the environment or human health, wasted resources, loss of credibility and sometimes, legal liability (Lombard and Kirchmer 2004). Data quality, however, fundamentally depends on the intended use of the data. To be meaningful, the data quality must meet the desired level of confidence for the purpose of the sampling program. As well, the sampling design and data quality should be able to perform over a wide range of possible outcomes.

As a general policy, surface water and sediment sampling programs conducted by AENV use accredited laboratories, although new and emerging substances may not be accredited yet. Laboratory accreditation, however, does not guarantee good data. Many other factors can influence data quality.

To ensure that good data are collected, all sampling programs should include a quality assurance plan. *Quality assurance* (QA) is a system of activities designed to make sure that the data meet defined standards of quality. It pertains to the overall management of the sampling program, and includes planning, documentation, training, consistency in collecting and handling samples, analyses, validation and reporting. An important part of QA is *quality control* (QC). Quality control refers to the technical activities used to reduce errors throughout the sampling program. These activities measure the performance of a process against defined standards to verify that the data meet the expected quality. Errors can occur in the field, laboratory or while handling the data. QC should include both internal and external measures. Internal QC is a set of measures undertaken by the project's own samplers and analysts. External QC involves people and laboratories outside of the project (USEPA 1996).

Table 1 shows how QA and QC differ.

Quality assurance	Quality control
• Anticipates problems before they occur	• Responds to observed problems
• Uses all available information to generate improvements	• Uses ongoing measurements to make decisions on the processes or products
• Is not tied to a specific quality standard	• Requires a pre-specified quality standard for comparability
• Is applicable mostly at the planning stage	• Is applicable mostly at the processing stage
• Is all-encompassing in its activities	• Is a set procedure that is a subset of quality assurance

Table 1. Comparison of quality assurance and quality control. *From Statistics Canada:* http://www.statcan.ca/english/edu/power/ch3/quality/quality.htm

The purpose of this document is to recommend QA/QC guidelines for all water and sediment sampling programs conducted by aquatic scientists in the Alberta government. To make sure that all data generated on various projects are reliable, quality assurance

and quality control must be included in sampling designs for all water quality monitoring projects; quality control data should be interpreted in project reports. This document is intended primarily for the staff of the Alberta government and their partners and consulting firms conducting studies for AENV. It would also be useful for anyone doing water quality studies.

Although several jurisdictions in North America have quality assurance guideline documents for surface water sampling, they vary considerably in methods and requirements. The information herein may differ somewhat from literature sources, but the guidelines were agreed upon by Alberta Environment staff and should be applied to Alberta government sampling programs.

These guidelines are intended for sampling of surface water and sediments only, including chemistry and a few biological variables, mainly fecal coliform bacteria and chlorophyll *a*. Other biological variables require different QA techniques, and likely different guidelines. This document does not address compliance monitoring for regulatory purposes, nor for groundwater monitoring, although the principles would be the same. A field sampling procedures manual, which includes field QC, is available (Alberta Environment 2006), and therefore specific field techniques to ensure quality data are not reported here, nor are those for certified analytical laboratories, which have their own QA/QC procedures. Appendix A provides information on the current state of quality assurance in Alberta Environment.

2. BASIC CONCEPTS OF QA/QC

Quality assurance is the overall management of a sampling program so that reliable and accurate data are produced. Variability occurs naturally in streams, lakes and rivers, but is also introduced during the collection and analysis of samples from these waters. For analytical results to be meaningful the total error contributed by all stages of sampling and analysis should be substantially less than the natural variability of the water or sediments being sampled. All of the following would apply to sediments as well as water.

Field quality assurance includes basic precautions that must be followed if variability (errors) in the data is to be minimized. The Alberta Environment field-sampling manual (Alberta Environment 2006) gives specific instructions to maintain consistency and ensure the staff are diligent while collecting, filtering, preserving and shipping samples. Quality control (QC) samples are used to evaluate whether the sampling and processing system is functioning properly, and whether measurement quality objectives have been met. Analytical labs have their own quality control procedures, but QC samples submitted from the field will provide an estimation of the total study error. If necessary, QC samples can be used to pinpoint sources of error, especially those from contamination. New sampling programs should incorporate rigorous QC measures until an acceptable level of data quality has been demonstrated. This is particularly important if the program objectives are to assess trends or investigate an impact on aquatic life or human health.

Measurement Quality Objectives and Data Quality Objectives

A practical distinction between error and uncertainty is that we can do something about error, but we have to live with uncertainty. Error has to do with the quality of measurements, while uncertainty has to do with what they represent. In practice we can identify sources of both, and both can be diminished. The level of effort depends on the amount of error we can tolerate. Error is how far out a particular measurement could be from the "truth" – it can be specified as a percentage of the true value, such as +/- 10% (Katznelson 1998). These are often referred to as Measurement Quality Objectives (MQOs). MQOs are specific units of measure, such as percent recovery (accuracy) and percent relative standard deviation (precision). They are usually listed in the same units as the real sample data, so they can be compared directly to QC sample results. They should be specified before a sampling program begins, and the QC data should be analyzed during the sampling program, so that problems that may arise can be corrected.

Data quality objectives (DQO) are statements that define the degree of confidence in conclusions from data produced from a sampling program. They are essentially steps to follow to make sure the data are adequate to address the environmental issue being studied. DQOs are especially important when an environmental decision is to be made, or for comparisons between reference and impacted sites. The USEPA has several guidance documents (e.g., USEPA 2000b) that detail the process of formulating DQOs.

Quality Control Indicators

Quality control indicators include precision, accuracy, representativeness, completeness, comparability and sensitivity. With these indicators, we can evaluate the sources of variability and increase confidence in the data produced. Each sampling program could have different levels of acceptability for these indicators, depending on the goals and data needs of the program. The following is derived from Cavanagh et al. (1997) and USEPA (1996).

Precision

Precision is the degree of similarity among measurements taken from two or more subsamples of a particular sample (splits), or from two or more repeated samples collected from water or sediment as close as possible in time and space (replicates). From these, precision of the entire program (laboratory + field) can be estimated. Close agreement among measurements means that sampling and analysis are giving consistent results under similar conditions. Imprecision is the result of inconsistent field techniques, laboratory analysis or both.

Replicates collected consecutively from a water body would include natural variability, and are inherently less precise than splits from a single sample, although this variation is expected to be low. Precision calculated from split samples would not include natural variability, although it could include variability from the splitting process. If possible, three rather than two samples (duplicates) should be collected or split because increasing the number of samples may increase precision.

Precision is calculated as the relative standard deviation (RSD) of the split or replicate data, with precision decreasing as the RSD increases. For duplicate samples, it is calculated as the relative percent difference between the two samples (*see Chapter 4)*. RSDs for the total program are often several times greater than those of the laboratory. It should be remembered that precision cannot be calculated from non-detect data, and for some variables, only replicates (not splits) can be done due to issues related to sampling methodology requirements (for example, for certain organics and trace level mercury).

Sediment samples are generally less precise than water samples, and therefore many project managers collect composite samples of three or five replicates. For environmental effects monitoring in sediments, Environment Canada (2005) states "the collection of replicate samples should be mandatory as part of the QA/QC requirements" and that a minimum of five replicates should be collected. If composite samples of sediments are collected, three replicate composite samples should be collected at the same site or one composite sample should be split into three portions (triplicate split). If replicates are done, is very important to make sure the sediments collected are similar in terms of depth, location, water flow (for streams) and other factors.

Accuracy

Accuracy is a measure of the agreement of a value of a variable in a sample with a known or "true" value. Random and systematic errors during sampling and analysis of samples reduce accuracy. Random errors refer to random variation or precision of the data, whereas systematic errors refer to bias, or values that are consistently higher or lower than the true value. Historically, the term *accuracy* has been used interchangeably with bias, but for the purpose of this guideline, accuracy includes precision as well as bias (Figure 1).

Accuracy is measured with standard reference materials with known values, either directly or as a spiked matrix sample. The quantity of a variable in a spiked sample is compared with that in a duplicate sample without the spike. It is usually expressed as percent recovery. A few variables, for example fecal coliform bacteria, cannot be tested for accuracy because standard reference materials are not available. In this case, split or replicate samples can be submitted to different labs. It should be remembered, however, that this does not give information on the true value of a substance, because it will not be known which lab is the more accurate, or if either of them are. Taking more samples does not reduce bias.

PRECISION, BIAS, AND ACCURACY
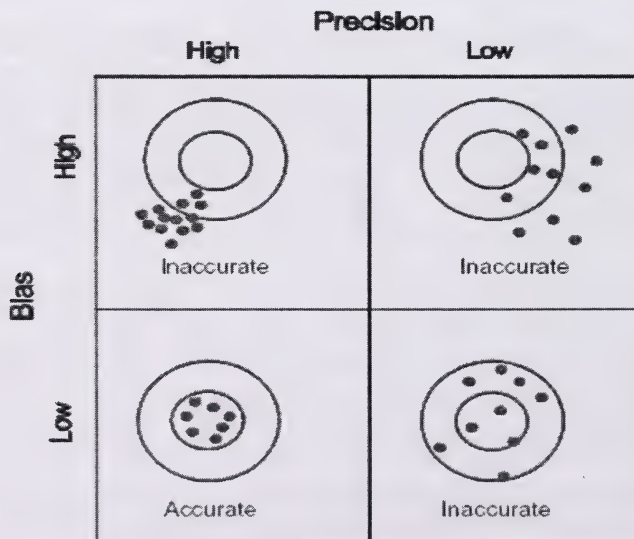
Figure 1. Accuracy, precision and bias in water analyses (from USEPA 1996).

Other Estimates of Quality Assurance

Representativeness is a qualitative term that expresses how well the data reflect the true environmental condition being sampled. It includes accuracy and precision, but also how well the sampling site reflects the water body under study. Variability in sampling should be minimized to ensure representativeness. For example, a sample collected immediately downstream of a tributary input on a river would likely not be representative of the river under study.

Comparability is a qualitative term that expresses how well one data set from a particular water body can be compared to another from the same water body. It compares sample collection and handling methods, analytical techniques and QA protocols. For example, if discrete or grab samples had been collected for a former study, it would not be appropriate to do composite samples for a new study if the purpose was to compare data from the two studies.

Completeness is a measure of the amount of valid data needed to assess the hypothesis of the program. Several statistical procedures are available to estimate the number of samples needed for a particular study design (*see Chapter 4 and USEPA 2000b*).

Sensitivity is the capability of a method or instrument to discriminate between measurement responses for different levels of the variable of interest. This generally refers to *detection limit* or to an instrument's measurement range. Further information on detection limit may be found in Chapter 3 and Appendices A and C.

Quality Control Samples

Quality control samples are collected in addition to the regular samples in the program design. Field QC samples should generally be submitted blind, that is, labeled so the lab would not know they are for QC. The following is compiled from Cavanagh et al. (1997), USEPA (1996), McQuaker (1999) and Mueller et al. (1997).

Blanks

A blank is water that is free of the analyte to be measured, and is prepared by the analytical lab. The preparation or source of the water will vary according to the analysis required. For example, VPP or volatile organics blank water is prepared from distilled - deionized water (with carbon trap) followed by boiling in organic free environment (Grant Prill, pers. comm.). Blanks can identify unsuspected contaminants from the water itself, improper cleaning procedures, preservatives, samplers, filters, travel, sampling techniques or air contaminants. In general, blanks should be analyte-free. The following is modified from McQuaker (1999):

- *Field blanks* are exposed to the sampling environment at the site and handled as a "real" sample. They provide information on contamination from all phases of sampling and analysis. They should be done routinely and at a reasonably high frequency. Field blanks should be prepared during or immediately after collecting

environmental samples at the sampling site. Other types of blanks are used to pinpoint problem areas.

- *Trip blanks* are bottles filled with blank water at the lab and transported to and from the field without being opened. For certain analyses, preservatives might be added at the lab before being sent to the field staff. They are used to check for contamination from the bottles, caps or preservatives. They are especially useful for volatile compounds, but can be used for any type of variable. Trip blanks can become contaminated when stored for long periods, and therefore should be used within one week.

- *Equipment blanks* are samples of blank water that has been used to rinse sampling equipment in the field. They are preserved and analyzed in the same manner as the samples.

- *Filtration blanks* are blank water samples that that have passed through the filtration apparatus in the same manner as the sample. These can pinpoint cross-contamination through inadequate cleaning techniques.

Reference Samples

Reference samples are used to document the bias and precision of the analytical process. The simplest type of reference sample is provided by a laboratory that is not involved in analysis of the real sample. The reference sample is prepared by adding a known quantity of the variable of interest to a given amount of pure water. Another type of reference sample is a Certified Reference Sample (CRS). It is a portion of a very large batch sample that was collected from one place at one time. It has been analyzed by a large number of independent labs, and therefore the distributing agency can provide a mean and confidence interval for the variable of concern. It is obtained from a scientific body such as the National Research Council. Although laboratories use reference samples for their own QC, reference samples should be submitted blind to the lab before each new sampling program begins. A submission of multiple samples of the same CRS yields the laboratory precision, as well as bias.

Spikes

Spiked samples are prepared by adding a pre-measured amount of a reference sample to a single water sample. When natural water (as opposed to blank water) is used, it is called a matrix spike. An aliquot of the same sample is left unspiked. A plain blank and a spiked blank should also be submitted to make sure the spikes were done properly. Accuracy is estimated by subtracting the amount of analyte in the unaltered split sample from the amount in the spiked sample. The amount left should equal the amount of the spike. *Recovery* is the percentage of the recovered amount of analyte. This reveals systematic errors (bias) in the analytical method. Bias can arise from matrix interference or analyte degradation. Spiked blanks submitted along with matrix spikes could help

assess interferences, which can affect recovery rates. Recoveries will vary with the analyte, matrix, batch and laboratory, especially for organics.

Replicates

Replicates are used to calculate precision. Two or more samples collected in quick succession from the same location in the water body will yield precision of the entire program. It also includes variability of the water body itself, but this should be minimal compared with other sources of variability. It is generally better to collect three replicates (triplicate) if the budget will allow it. As well, five replicates should be collected for variables such as fecal coliform bacteria, which have a high natural variability. Replicate results that are non-detects cannot be used to estimate precision, so the project manager should select samples or measurements likely to yield positive results.

True Splits

A true split is a single sample that has been divided into two or more portions. It is better to submit three portions if possible (triplicate split), to detect outliers and reduce variance, especially if the concentrations are at trace levels. If the sample is split in the lab, these can check laboratory precision. If done in the field, these yield field + lab precision. When splits are done in the lab and field (i.e., one of the field splits is split again in the lab), the lab precision can be subtracted from the total precision to yield field precision. Sometimes splits are sent to different laboratories for analysis. However, there is no way to determine which laboratory's results are correct. Instead, the project manager should submit reference materials or do a performance test before engaging the lab. Non-detect results cannot be used to assess precision.

Data Management

Data management addresses the path of data from the field and laboratory to its final use and archiving. Errors can occur anywhere along this path. Laboratories verify the data before they are reported to the project manager. This includes an evaluation of their QC results. Project managers then verify and validate the data. According to USEPA (2002b), data *verification* is the process of evaluating the completeness, correctness and conformance of a specific data set against method, procedural or contractual requirements. Data *validation* is an analyte- and sample-specific process to determine the analytical quality of a particular data set. The goal of data validation is to evaluate whether the data quality objectives established during program design have been met. Validation can identify some but not all of the sampling error associated with a data set. For the purposes of this guideline, verification and validation as a whole is called data validation. *See Appendix D for further information on data validation.*

3. SAMPLING PROGRAM DESIGN

For many environmental studies, the planning, design and documentation of the sampling program is the most neglected area, with insufficient time allocated to it. But sampling program design is fundamental to quality assurance. A study must be carefully planned so that the resulting data are technically and scientifically sound, and to avoid waste of resources and time. As well, the design should include feedback loops to correct problems and make sure that the data are good enough for the objectives of the study. The key quality assurance question is, *"How much error can you tolerate for the purpose of the study"?*

An important aspect of any water or sediment sampling program is whether the data are representative of the water body being sampled. This is addressed through the sampling design. If a sampling design results in the collection of non-representative data, even the highest quality laboratory data are not valid for answering the problem at hand (USEPA 2002c).

For monitoring by partners, volunteers or consultants for AENV, the program design should be reviewed by professional staff of Alberta Environment. This will help ensure consistency throughout data collection in the province. It is important, however, that the reasons for this are explained, especially for volunteers. Appendix B provides further information on quality assurance elements in sampling programs.

Steps in Program Design

Systematic planning is based on the scientific method. It includes such concepts as objectivity of approach and acceptability of results, and ensures that the level of detail in planning is commensurate with the importance and intended use of the data and available resources. The steps listed below are designed to make sure the project manager thinks the sampling design through. The steps can be adapted to suit each manager, as long as the general principles are followed. Much of the following was taken from Cavanagh et al. (1997), USEPA (2000b) and Lombard and Kirchmer (2004).

1. State the Problem

The first step is to define the problem that has initiated the study. This step should include background on the water body and a summary of existing information. It could include a literature search or an examination of past or ongoing similar studies on other water bodies. If data on the water body under study are available, the QC performance of sampling and analytical methods from them should be assessed. It is critical to develop an accurate conceptual model of the environmental problem or issue, because this will serve as the basis for all subsequent inputs and decisions. This could include potential sources of pollutants, what effects there might be on human health or aquatic life and the type of media that should be looked at. Important water quality guidelines should be listed.

The key question or objective of the study should be stated. For example, "Are the types and concentrations of pesticides increasing in the Bow River"? Objectives should be specific, concise, understandable and meaningful. The actions to address the issue should also be stated. What would be done if it were found that certain pesticide levels are increasing in the Bow River? The actions might include finding the source of the pesticides and correcting the problem, conducting a more intensive study or doing nothing.

Outputs for this step are:
- Background and a narrative of the problem
- Possible effects on uses of water
- Specific objective of study

2. Identify the Information Needs

For this step, the kinds of information needed for the study should be determined. For example, do you use existing data, or do you go out and collect more? What do you need to measure to address the question, and are appropriate analytical methods available? If flow volumes of the river or tributaries are needed, is there an appropriate Water Survey of Canada (WSC) or provincial hydrometric station within the watershed, or will it be necessary to set up a new station? If modeling of the water body is a possibility, then a list of input variables is required. As well, it is important to make sure that if new data will be compared with existing data, sampling and analytical methods are similar. If the study involves an environmental decision, you will need to determine the action level. For example, should you use a water or sediment quality guideline? Or should you do a risk assessment? If the decision will be made relative to background or reference conditions, then you should determine what constitutes background. You should also identify analytical methods that have appropriate detection limits, for example, ten times below guideline levels. If it is a new study, or the location is unfamiliar to you, you should visit the site(s) and record field observations.

Outputs for this step are:
- General environmental characteristics to be measured to address the issue
- A preliminary documentation of field observations
- A list of information that may be applicable to uses of the data in future, such as inputs to models
- Confirmation that sampling and analytical methods exist and meet the detection limit requirements

3. Define the Boundaries of the Study

In this step, the spatial and temporal features of the water body to be sampled should be specified. You would define the spatial boundaries - for example, the stretch of river from 20 km above the city of Red Deer to 50 km downstream - and then specify where samples will be collected. Temporal boundaries include the time frame for the study, and when the samples will be taken. Land use constraints for sampling would also be

included here, as would seasonal issues such as spring runoff or stability of the substance under study. You would also describe limitations imposed on the schedule, such as equipment needs, seasonal conditions, time needed for sampling and resources available.

Outputs for this step are:
- Detailed description of the characteristics that define the water body to be sampled
- Spatial boundaries and approximate sampling sites.
- Time frame for the study.
- List of practical constraints that may interfere with data collection

4. Develop a Testable Hypothesis for Assessment Studies

In this step, you should develop a hypothesis to be tested statistically. If the purpose of the study is to assess impacts, analyze trends or make an environmental decision, you will need data that are statistically adequate to do this. Even when the sampling and analytical methods are unbiased, the data are subject to random and systematic errors. You should establish the statistical requirements before sampling begins, not after the data are collected.

For an assessment type of study, you can use hypothesis testing to choose between a baseline condition of the environment and an alternative condition; for example, between upstream and downstream of an impact, or a temporal trend or no trend. In statistics, the baseline condition is called the null hypothesis (H_O) and the opposite condition the alternative hypothesis (H_A). For example:

H_O: The mean concentration of aluminum in the North Saskatchewan River downstream of Edmonton is the same as that upstream.

H_A: The mean concentration of aluminum in the North Saskatchewan River downstream of Edmonton is different (or higher) than that upstream.

The test can show whether 1) there is insufficient evidence to indicate that there is no difference between the baseline condition (or null hypothesis) and the alternative condition (and therefore you accept that the baseline condition is true), or 2) that the baseline condition is false, and therefore the alternative condition is true. In this way, the burden of proof is on rejecting the null hypothesis, for which you need overwhelming evidence to support this. It is critical to understand that the selection of the baseline condition is important to the outcome of the decision process of the analysis. A false rejection error can lead to the conclusion that the baseline condition is false when it is really true (Table 2). The opposite conclusion can also result, that is accepting the baseline condition as true when it is really false.

Decision based on Sample Data	True Condition	
	Baseline is True	Alternative is True
Baseline (Ho) is true	Correct Decision	*Decision Error (False Acceptance)*
Alternative (HA) is true	*Decision Error (False Rejection)*	Correct Decision

Table 2. False acceptance and false rejection decisions. *From USEPA (2000b).*

Sometimes the purpose of a sampling program is for an environmental decision, and requires that an unambiguous testable hypothesis is constructed. This is done as a theoretical *"If...then..."* statement. For example,

"If the concentration of dimethoate in any sample from the Red Deer River is found at one-half of the Canadian Water Quality Guideline for the protection of aquatic life or higher, then sources will be investigated. If the concentration of dimethoate is less than this, sources will not be investigated, although monitoring will continue."

Such hypotheses could be made for each of the pesticides being examined, for example, in the Red Deer River, although usually only one or two are of concern.

You should also confirm that the method detection limit for each of the variables is adequate to test the hypothesis. In the above example, the detection limit from most labs is well below the guideline concentration. Detection limits are defined specific to an intended purpose. The detection limit should be at a level that provides for a high probability of positive identification and presence in the matrix and a low probability of false identification.

Output for this step:

- Hypothesis or decision statement to meet objective of study

5. Set Data Quality Objectives and Measurement Quality Objectives

Criteria for accepting or rejecting data based on QC samples should be set. For certain studies, only measurement quality objectives (MQOs) are needed, whereas for others, both data quality objectives (DQOs) and MQOs should be developed. In general, trend assessment, impact studies and environmental decisions require DQOs, although baseline studies may be used for trend assessment, and should be well documented, with good QA/QC procedures.

Data Quality Objectives. The single most difficult aspect of any sampling program is natural variability in space and time. A major goal of the sampling design is to reduce variability and hence decision errors while having a cost-effective sampling program. If no regard is given to natural or study variability, then it will be impossible to determine impacts or changes in water or sediment quality.

A Data Quality Objective (DQO) is the confidence level for the data you wish to collect. This is generally set from previously collected data. If, after analyzing the previous data, you find that the confidence level (variability) for a critical substance is very high, you would need to increase the number of samples, or stratify the samples so they better represent the water body. If no data are available, it may be cost effective to conduct a limited field investigation to get preliminary estimates of variance to determine the number of samples required. Existing data should also be examined for indications of analytical problems, such as poor detection limits or those that change over the course of the study.

For some decisions, a defensible conclusion can be drawn from relatively small sample sizes or data that are relatively imprecise. Conversely, if the consequences of decision errors are severe (for example possible harm to aquatic life), you will need to have more control over sampling design and measurement error. You will have to balance the consequences of decision errors against the cost of limiting the possibility of these errors. The potential consequences of each decision error should be evaluated, as well as the magnitude of the decision error. The consequences of a decision error when the true parameter value is only 10% above a water quality guideline level could be very different than if the value were 10 times above this level.

You will also need to set the statistical significance level for accepting or rejecting the null hypothesis for the variables of concern. These are set arbitrarily. In general, probabilities for false rejection errors are often set at 0.05 for environmental studies, meaning that there is a one in 20 chance that the null hypothesis will be rejected when in fact there is no difference. If the water body is of very high value, you might want to increase the possibility of making this type of error, but decrease the possibility of accepting the null hypothesis when there might be an actual impact. It is obvious that actual probabilities will depend on the study requirements.

Statistical software is available to calculate the number of samples needed to satisfy data quality objectives. The U.S. National Water Monitoring Council has software, called DQO-PRO, to calculate DQOs for the number of samples to be collected to achieve required confidence levels, or the confidence levels for the number of samples you wish to collect:

http://www.envirofacs.org/dqopro.htm

Measurement Quality Objectives. Measurement quality objectives (MQOs) also specify how good the data must be to meet the objectives of the project. MQOs should be used for all water and sediment studies. They are based on the quality indicators precision, accuracy and sensitivity. MQOs are usually expressed in the same units used for reporting QC sample results. This facilitates data validation, because the QC sample results can be compared directly to the previously established MQOs to determine whether the sample data are acceptable. For example, for a study designed to assess metals in water samples, the following MQOs were established (Lombard and Kirchmer 2004):

- Lab quality control samples: 85% - 115% recovery
- Split or replicate sample analyses: less than or equal to 20% relative percent difference
- Matrix spike recoveries: 75% - 125%
- Duplicate matrix spikes: less than or equal to 20% relative percent difference.

For some variables, accuracy cannot be determined because no standard reference materials are available. For these, MQOs may be limited to precision of replicate or split samples. For field measurements, it is important to make sure that instruments are calibrated and checked frequently. For these, MQOs can be expressed as the maximum deviation allowed for calibration checks.

Alberta Environment is beginning the process of developing measurement quality objectives. The following MQOs for phosphorus and chlorophyll *a* in lake water are preliminary only (Table 3). These are based on lake data for 1996-2004.

Statistic	Phosphorus		Chlorophyll a
	Total P	Dissolved P	
Operational range	1 - 200 ug/L	1 - 200 ug/L	0.02 - 500 ug/L
Detection limit	1 μg/L	1 μg/L	0.02 ug/L
Accuracy (recovery)	97 +/- 8%	97 +/- 8%	N/A
Precision:			
- Duplicate splits	Not more than 4% difference	Not more than 3% difference	Not more than 2% difference
- Triplicate splits	Not more than 6% RSD	Not more than 2% RSD	Not more than 3% RSD
Detection limit: the minimum concentration with 99% confidence of being greater than zero. RSD: Relative standard deviation N/A: Not available			

Table 3. Preliminary MQOs for phosphorus and chlorophyll *a* in lake water.

MQOs for blanks should be established as well. For British Columbia, Cavanagh et al. (1997) states that not more than 5% of blanks for a study should exceed the MDL; blanks should not exceed 10% of environmental levels; levels in blanks should be less than 10% of relevant water or sediment quality guidelines.

MQOs for Phosphorus in Blanks in Alberta Water Sampling Programs

An overview of phosphorus data for Alberta revealed that approximately 22% of the field and trip blank TP data were above the detection limit (total number of samples 1996-2004 = 381); for TDP, 13% were above the detection limit (n = 328). The detection limit for these samples varied from 0.001 to 0.003 mg/L, depending on the laboratory. The vast majority of the data above the detection limit were at or very near the detection limit; very few approached ambient levels, and these tended to be in the older data. **A tentative MQO for phosphorus in blanks in Alberta sampling programs is that values in any blank should not exceed twice the detection limit.** For the Alberta data set, 6% of blank samples for TP and TDP would not have met the MQO.

- If a blank sample value for any variable of concern is observed to be above the analytical detection limit, even within the above MQO, additional blanks should be submitted on the next sampling occasion.

- If any blank exceeds the above MQO, a thorough investigation of possible sources of contamination should be done.

MQOs for sensitivity should be expressed as the lowest concentrations of interest or detection limit, usually 10 times lower than the water quality objectives for those variables. For some variables such as pH, however, it would not be meaningful to specify a lowest concentration of interest. Table 4 gives examples of MQOs for a few water quality characteristics for streams in the State of Washington (Hallock and Ehinger 2003). *Note that this is an example, and may not pertain to the Alberta situation.*

Analyte	**Accuracy (deviation from true value)**	**Precision (% relative standard deviation)**	**MDL (units)**
Total Nitrogen	20%	7%	0.025 mg/L
Ammonia -N	20%	7%	0.01 mg/L
Nitrite-Nitrate-N	20%	7%	0.01 mg/L
Total Phos.	20%	7%	0.01 mg/L
Sol. Reactive P	20%	7%	0.003 mg/L
TSS	20%	7%	1 mg/L
Turbidity	20%	7%	0.5 NTU
Fecal coliform bacteria	Not applicable	28%	1 colony/100 mL

Table 4. Example of measurement quality objectives for a hypothetical stream study in the State of Washington. *From Hallock and Ehinger (2003).*

Outputs for this step include:
- Assessment of previous data to establish variance
- Statistical significance level for rejecting null hypothesis, if required.
- The number of samples needed to ensure confidence in the data.
- A table summarizing MQOs for lab and field measurements. These should be compared with QC data as they are collected, so that adjustments or investigations can be made if a problem arises.

6. Finalize the Study Design

The design of the study will depend on whether it is an assessment or a monitoring study. Assessment studies are generally more complex, and require more detail in the monitoring plan. They require a statistical design, whereas a monitoring study requires a judgmental design.

The power of the statistical design can sometimes be improved by stratifying the samples. This is done by dividing the target population into strata that are relatively homogeneous. The strata are usually physically based, such as above and below an input that would affect water quality, or time based, such as seasonal. If the study involves sampling upstream and downstream of an impact, much of the natural variability can be compensated for by paired observations – that is, by sampling the two sites at similar times. The advantages of stratification are that it reduces complexity and variability and improves sampling efficiency. Disadvantages include difficulty in determining the basis for selecting strata and increase in cost because more samples may be needed, especially if the strata selected are not homogeneous.

A data collection design should include the following elements. Example spreadsheets for the design of two sampling programs are included in Appendix C:

- Sample locations (with GPS), and how they were selected
- Number of samples
- Sample type, for example composite or grab
- List of variables for each site
- Type and number of quality control samples (see below)
- General collection techniques or refer to sampling SOP manual
- Timing of collection, handling and analysis
- Analytical laboratories and methods for specialized analyses
- Statistical tests needed to analyze the data
- Type of personnel and manpower requirements
- Costs for sampling and analysis

Quality Control Samples

The number of QC samples that should be collected relative to the total number of samples is presented in Table 5. This is a general guideline only that should be varied depending on the program objectives. For example, you would collect more QC samples

when investigating a problem that could possibly harm aquatic life or human health (up to 35% of the analytical budget, Cavanagh et al. 1997), and fewer when routine samples are collected to monitor water bodies without major issues, such as recreational lakes (~5%). Regular monitoring programs (LTRN, trend assessment, etc.) would need somewhere in between these amounts. In general, when the substances of concern are at trace levels, more blanks should be collected; if concentrations are high, more replicates/splits should be collected.

	Frequency*	Quality Control Checks	Comments
Blanks			
Field	1/10	Contamination from bottles, collection methods, atmosphere, preservatives.	Treat as real sample.
Trip	1 per trip	Contamination from bottles, caps, preservatives	Not opened in field.
Equipment	1/Every fifth site	Carry-over between sites	When carry-over possible.
Filtration	As needed	Contamination from filters or filtration equipment	When contamination is suspected.
Spikes	1/20	Bias	Only for variables of concern
Replicates *or* Splits	1/10	Lab and field precision	Only for variables known to be above detection limit
Reference	Once at beginning of season or new project	Accuracy	Performance test before sampling begins

Table 5. General guidelines for types of quality control samples and their frequency of collection. *1/10 = 1 QC sample for every 10 regular samples. *Compiled from Cavanagh et al. (1997), Mueller et al. (1997) and AENV (2006).*

Examples of Study Design

The following examples are intended to show project managers the above steps in designing a sampling program. These examples are very simple, and you should remember that more complex sampling programs, especially those requiring biological components, several media and more water bodies, would be more difficult to design.

Example 1. Monitoring Study

Step 1. State the Problem

Lake L is a scenic, relatively deep lake in central Alberta. It is situated x km northwest of Edmonton. The lake has about 300 cottages in several subdivisions. Cottagers on the lake have become increasingly worried that water quality is deteriorating. They complain of murky green water, weeds and noxious smells as plant material in the lake decomposes, which inhibits recreational use. The county council for the Lake L area decided to ask Alberta Environment to assess water quality in the lake and determine causes of the problem. Lake L has not been sampled previously.

The key questions for this study are:

- What is the trophic status of Lake L?
- What are the major nutrient sources and their contribution to water quality in the lake?

Step 2. Identify Information Needs

The study will require new water quality data to be collected, including nutrients, chlorophyll *a*, field variables (D.O., pH, conductivity, temperature, and Secchi depth), routines, water levels and estimated inflows. Although some of these characteristics are not necessary to address the key questions specifically, they are part of a good baseline study. Preliminary field observations indicated that, in addition to the cottages and a resort camp, agriculture is present in the watershed. The sampling protocol and analytical methods used will be similar to that of other Alberta lake studies (see AENV 2006). A nutrient "desk" budget will also be prepared. For this, it will be necessary to obtain information on the number of lake users, sewage input coefficients and coefficients for nutrient loading in precipitation and surface runoff.

Step 3. Define Study Boundaries

Lake L has an area of 12 km^2, a mean depth of 7 m and a maximum depth of 18 m. The watershed area is 250 km^2. The lake will be sampled mainly during open-water for two years. The lake has a public boat launch, and access can be attained via good all-weather roads.

Step 4. Develop a Testable Hypothesis

This study does not require a testable hypothesis.

Step 5. Set MQOs

Measurement quality objectives were set as follows: Precision on triplicate split samples, 15% relative standard deviation; no analytes in blanks. Detection limits are adequate for this study and the chosen lab.

Step 6. Finalize Study Design

The water in Lake L will be sampled as whole-lake composites (15 random sites at depths of 3 m or more), every three weeks for two years during May-early October. Standard operating procedures for lake sampling will be used. Field profile measurements will be made at a designated location each trip (the deepest part of the lake), and one set of profile samples will be collected during winter each year. In addition, profile samples (2 m intervals) for phosphorus and nitrogen will be collected at the deepest site late May – early September to provide information on internal loading.

Quality control samples will include one field blank each year, and a triplicate split on one composite sample each year. Nutrients and routines only will be analyzed in the blank, and TP, TKN, chlorophyll a and TSS on splits. Triplicate field variables will be done once during the sampling program.

The sampling program will result in 20 composite samples and 16 profile samples, plus 6 QC samples (4 replicate and 2 blanks). Analytical costs are about $xxx.

Each sampling trip will require 2 people, for a total manpower requirement of 80 man-days.

Example 2: Assessment Study

Step 1. State the Problem

People using River B for recreation are seeing a greatly increased growth of algae on the bottom of the river in downstream reaches, but not upstream. They want to know what is happening. The river drains agricultural land in lower reaches, but is relatively unimpacted in upstream reaches. The river had been sampled fifteen years previously for a variety of substances at two sites, Site U and Site D. This study found that the mean concentration of total phosphorus was 0.029mg/L at Site U, the upstream site, and 0.103 mg/L at Site D, the downstream site. An increase in nutrient concentrations may help explain the increase in benthic algal growth.

The key questions for the new study are:

- Have the concentrations and loads of nutrients at Site D increased over the 15-year period?
- Are present flow-weighted mean concentrations of nutrients and amounts of benthic chlorophyll higher at Site D than at Site U?

Step 2. Identify Information Needs:

Data from the previous study will be used, but new data will be collected as well. The variables to be monitored include TP, TDP, TKN, NO2+NO3, ammonia-N, fecal coliform bacteria, routine variables (including TSS), benthic chlorophyll a and flow. Although some of these variables are being collected as adjunct information to the purpose of the study, the data may prove useful in future studies. A WSC station is present near the mouth of the river to obtain flow volume for Station D; it will be necessary to put a new flow gauge at Station U, and develop a rating curve. From this, flows can be estimated for the previous study by relating this information to the WSC gauge and developing runoff coefficients. Preliminary field observations suggest that several new cattle operations are present in the lower watershed, and extensive logging has occurred throughout the watershed. The previous sites and similar analytical and sampling techniques will be used. Detection limits are adequate for this assessment.

Step 3. Define the Study Boundaries:

River B has a mean annual flow volume of about 2000 m^3. It flows eastward from the foothills, and joins River C 400 km downstream. Peak flows occur in May and again in July. River B will be sampled above agricultural operations at Site U, the site sampled previously (lats and longs), and 200 km downstream at Site D (lats and longs). Both sites have good access.

Step 4. Develop a Testable Hypothesis:

The key questions and their hypotheses for this study are:

- Have the flow-weighted mean concentrations and loads of nutrients at Site D increased over the 15-year period?

  H$_O$: *Flow-wt mean concentrations and loads of total phosphorus (TDP, TKN, etc,) have not increased at Site D in River B between 1990 and 2006.*

  H$_A$: *Flow-wt mean concentrations and loads of total phosphorus (TDP, TKN, etc,) have increased at Site D between 1990 and 2006.*

- Are present flow-weighted mean concentrations of nutrients and amounts of benthic chlorophyll higher at Site D than at Site U?

  H$_O$: *Present concentrations of total phosphorus (TDP, TKN, etc.) and chlorophyll are the same at Sites D and U.*

  H$_A$: *Present concentrations of total phosphorus (TDP, TKN, etc.) and chlorophyll are higher at Site D than at Site U.*

Step 5. Set DQOs and MQOs.

In the first study, a total of 18 samples were collected at Site U, and 13 at Site D. An assessment of the data from the first study found that concentrations of some of these substances varied considerably over the year: at Site U, range in TP = 0.002 – 0.140 mg/L; Site D, range in TP = 0.004 – 0.361 mg/L. This variability would likely mask any differences between the time periods or sites. Therefore, for the new study, the sample size will be increased, and the samples collected on the same day, with Site U sampled first. The data from the new study can be treated as paired observations. Statistical software runs suggested that about 30 samples would be needed at Site D to reduce the variance and improve confidence limits.

Measurement quality objectives were set as follows: Precision on triplicate samples, 15% relative standard deviation; accuracy from spikes, +/-10% deviation from the true value; values in blanks should be no more than twice the MDL for any variable. Detection limits are adequate for this study from the chosen lab.

Step 6. Finalize Study Design.

The first year of the study will focus on obtaining flow data for Station U. A water level recorder will be set up during spring, and manual flow gauging will be done twice per week during spring runoff and once per week during the rest of the summer. For the following two years, both flow and water quality information will be collected.

Water quality samples will be collected as three-point composites across the river at both sites according to the sampling SOP manual (AENV 2006). Field variables will be done at the center site (D.O., cond., pH, temp). Water chemistry and fecal coliform bacteria samples will be collected twice-monthly March 15 – October 15, resulting in 15 samples at each site per year, for a total of 60 samples. Three benthic chlorophyll a samples will be collected once in July, August and early September at each site each year. No winter samples will be collected. Table xx shows the variables, detection limits and timing for this program.

In addition, one field blank and one triplicate split will be collected at both sites each year and one nutrient spike will be done at Site U early in the season each year. Blanks will be analyzed for all variables including fecal coliform bacteria, while the triplicates will be analyzed for TP, TKN, TSS and fecal coliform bacteria. Triplicate sets of benthic chlorophylls will be collected once each year at Site D. Samples will be submitted to Maxxam Labs and other labs. The total number of QC samples is 14 plus 6 benthic chlorophyll.

The data will be analyzed for equality of variance and normality, and depending on the outcome, a t-test or Mann-Whitney test will be done for the first hypothesis, and a paired t-test or Wilcoxon's test will be performed for the second hypothesis. The probability of rejecting the null hypothesis is set at 0.05 for all tests.

Approximately 300 man-days will be needed to complete the study.

4. DATA QUALITY ASSESSMENT

The scientific and statistical evaluation of quality assurance data is called *data quality assessment*. Its purpose is primarily to determine whether the data collected during a sampling program are the right type, quality and quantity to support their intended use. Data quality does not exist in a vacuum; one must know in what context a data set is to be used to be able to judge whether or not the data set is adequate (USEPA 2000a).

As soon as the data are received, the project manager should immediately screen it for typos, unrealistic results, or obvious errors. Then, the project manager should:

- Compare the electronic field data with the original field data sheets.
- Examine field reports to check holding times for samples with sensitive variables, and flag any that exceed recommended times (this can be done electronically).
- Look for obvious deviations in the data, for example, when a result greatly exceeds or is greatly less than that of previous data.
- Check for illogical situations, such as the concentration of dissolved phosphorus exceeding that of total phosphorus.

After this initial assessment, the project manager should assess precision, accuracy and blanks, and compare the results with measurement quality objectives (*see below*). This should not be left until the data for the whole program are available, but it should be done during the sampling program to be able to correct problems as they arise.

The project manager should investigate the flagged results further. This may require discussion with the laboratory or field staff. Then, the results are accepted, deemed not conclusive or rejected. Generally the accepted and non-conclusive data are filed, while the rejected data are deleted from the database or flagged as invalid. According to Am. Public Health Assn (2005), theoretically no result should be rejected, because it may indicate either an analytical problem that casts doubt on all results, or the presence of a true variant in the distribution of the data. In practice, reject the result of any analysis in which a *known error* has occurred, for example a documented case of contamination. But data should not be rejected arbitrarily, because they may indicate areas in the water body with actual contamination, or areas without the variable of interest.

Quality Control Sample Results

Quality control sample data will allow the project manager to assess precision (splits), accuracy (spikes) and contamination (blanks). This will provide an estimate of the total uncertainty in the data set, in other words, the variability associated with sampling and analysis. When the QC data exceed measurement quality objectives, all associated sample values should be viewed with caution and the problem discussed in the final report. It is up to the discretion of the project manager whether to interpret these data, but they should be included in the report. The authors should provide rationale for

excluding data from interpretation. (Govt. BC 1998). Project managers should track corrections or changes so that laboratory performance can be assessed after the completion of the sampling year.

Contamination

Data from blanks should be examined to assess contamination and therefore bias. Levels of any substance found in blanks should be compared with MQOs for blanks. When blanks have measurable levels of particular substances, it is likely that the regular samples are biased. The objectives of the study should be considered when deciding how to treat the real data when contamination has occurred. If the objective is to detect minute changes in variable concentrations, then even low levels of contamination reduce the ability to interpret the data with confidence. But for certain monitoring studies, low levels of contamination may not be significant. In some cases, it may be valid to subtract blank values from the real data, especially if the real data are not particularly variable, or for certain trace level investigations. But whenever blanks exceed the MQOs, the source of contamination should be addressed to eliminate it in the future (Govt. BC 1998).

INVESTIGATION OF CONTAMINATION DURING SAMPLING OF A RIVER

Six blanks were collected during a water quality survey of River M. The following results (in ug/L) were obtained for three metals:

Dissolved Lead: 0.272 0.428 0.096 0.093 <0.01 <0.01

Total recoverable Copper: 0.08 0.27 1.19 0.11 0.35 0.12

Dissolved Zinc: <0.2 nd 3.86 0.36 0.32 0.25

Although none of these values exceeded water quality guidelines, lead concentrations in some of the blanks were higher than those in the ambient samples, and many of the blank data exceeded MQOs. The project manager decided to do an extensive investigation by submitting blanks from throughout the sample collection and processing. It was found that the filters used for dissolved compounds were contaminated, and that in some instances, preservatives may have been contaminated. As a result, the old filters were discarded, and a new supplier was found. The new filters were checked thoroughly and found to be ok. The cause of the preservative contamination was not discovered, but a new batch of preservatives was ordered and checked, and found to be clean. Ambient data for these metals were not used for interpretation of water quality conditions, although they were reported in appendices and an explanation was provided in the quality assurance section of the final report. *Data from McEachern and Noton (2002).*

<u>Precision</u>

Data from split or replicate samples can provide an estimate of precision. If duplicates were collected, calculate relative percent difference as:

$$(\text{Value A} - \text{Value B})/ \frac{(A+B)}{2} * 100$$

When three or more replicates are available, calculate percent relative standard deviation (RSD) as:

$$s/mean*100$$

where s = standard deviation of the mean.

Precision is influenced by how close the analytical value is to the MDL. Precision decreases rapidly the closer the values are to the MDL (or, RSD increases). Therefore, analytical values should be at least five times greater than the MDL. In general, data should be viewed with caution when the relative % difference for duplicates is greater than 25%, and the RSD for triplicates 18%. If replicate data are above these values, the source of variability should be assessed. This would likely mean submitting additional QC samples to assess sources of contamination, especially blanks (Govt. of BC 1998).

ASSESSMENT OF PRECISION DURING A RIVER STUDY

A water sampling program was initiated on River B. Samples were collected weekly, and over the sampling period, three sets of triplicate QC samples were collected. These were done as separate samples, collected no more than one minute apart from the river (replicates). The following data for total phosphorus (in mg/L) were obtained:

Set 1: 0.023, 0.026, 0.024
Set 2: 0.060, 0.072, 0.068
Set 3: 0.023, 0.042, 0.031

Descriptive Statistics:

Set 1: mean = 0.024, s = 0.002, RSD = 6%
Set 2: mean = 0.067, s = 0.006, RSD = 9%
Set 3: mean = 0.032, s = 0.042, RSD = 30%

Precision on Sets 1 and 2 is good, but that of Set 3 is poor. It exceeded the MQO of 20% established for the sampling design. The reason for this was not obvious, as turbidity in the river was fairly low. Blanks were relatively clean, so the project manager decided to collect more blanks and replicates the next time the river was sampled, and also include a phosphorus spike. The manager kept the data for that date, but did not include it in statistical assessments of the river data. Further assessment of precision in subsequent data showed that the problem did not recur, and it was concluded that contamination had occurred either in the field or lab.

Accuracy

Reference or spiked samples provide an estimate of accuracy. Standard (or certified) reference samples can be submitted to one or more proposed laboratories before the start of the sampling program to assess the labs' performance. This could be done annually and the results used for many sampling programs. Levels in the reference samples should be within the range of values expected in the water body being sampled. Certified reference samples come with a range of acceptable values set as 3 standard deviations around the mean, and the reference sample provider includes these.

Field spiked samples can also assess accuracy. A natural sample should be split, with the reference material added to one half and the other half left unspiked. Then, the concentration in the unspiked sample is subtracted from the concentration in the spiked sample, and the result compared with known values for the spike. Accuracy can be expressed as a percentage by dividing the analyzed value by the true concentration, and multiplying by 100%. MQOs should be established for accuracy in field spikes, such as +/- 10%. If a laboratory's results are higher or lower than this, you should consider that the real samples are similarly biased. The lab should be contacted to address the problem (Cavanagh et al. 1997).

A consistently biased data set will not affect nonparametric trend analysis. But if the bias changes over the course of data collection (such as a change in analytical laboratories or detection limits), the statistical analysis will be compromised. If a laboratory change is expected to occur, old and new procedures (including paired samples) should be overlapped for several months to assess bias.

ASSESSMENT OF ACCURACY IN WATER SAMPLES FOR TWO LABORATORIES

Analytical performance of trace metals was tested in two Alberta laboratories. River samples containing spikes of certified reference material of known concentration were provided to the labs. The reference materials were diluted according to instructions given by the supplier, except that one-half of the suggested concentration was used to better represent the concentration ranges typically found in Alberta surface water. The samples submitted to each lab were as follows:

- Blank (reagent grade water preserved with HNO3);
- Spiked blank water;
- Low concentration river water;
- Spiked low concentration river water;
- Higher concentration river water;
- Spiked higher concentration river water.

The unspiked values for each concentration were subtracted from the concentrations in the spiked samples. The following examples of results (in ug/L) were obtained:

		Arsenic	Copper	Lead
Blank	Lab 1	<0.2*	<0.2*	<0.1*
	Lab 2	<0.2*	0.5 (MDL=0.2)	<0.3*
Low concentration spike	Lab 1	19	38	37
	Lab 2	**12**	41	40
High concentration spike	Lab 1	19	35	37
	Lab 2	23	39	39
Acceptance range		**16-24**	**34-46**	**34-46**

*method detection limit

All of the results are within the acceptance criteria except for arsenic in the low concentration river sample at Lab 2. The lab will be contacted to assess this problem. One blank was positive for an analyte at Lab 2. But since the concentration is much lower than that in the other samples, this level is acceptable. *From McDonald and LeClair (2006).*

Non-Detect Data

Any laboratory used in environmental studies should be able to routinely report values at or below the lowest concentration of interest for the study. This generally means that detection limits should be 10 times lower than the water or sediment quality guideline for the substance. McDonald and LeClair (2006) assessed trace metals data for 2003 from the Alberta Long-term River Network. Its purpose was to determine how comparable the data from two analytical labs were. For several variables, the detection limits at the two labs differed. Water quality guidelines were used to test the utility of reported concentrations. A greater percentage of acceptable detection limits (less than 10% of the guideline value) were present at one lab compared with the other. If a particular

substance is of concern, for example mercury, more sensitive analytical techniques could be used.

Environmental scientists tend to be unsatisfied with data that are below detection, because of information loss. Censored data greatly reduce the power of statistical tests and create confusion about how to treat the censored results. Three approaches for handling non-detect data are generally used: 1) substitute the reported detection limit for all non-detects, 2) substitute one-half of the detection limit for non-detects, and 3) substitute a value of zero for non-detects. However, substitution of some fraction of the MDL for non-detects may result in inaccurate statistics, misleading regression models and inaccurate environmental decisions. Numerous studies have found that substituting one-half of the MDL is inferior to any other method, unless the proportion of missing data is small (10% or less). The problem worsens for long-term trend analysis when multiple MDLs are in the data set (Helsel 2005).

Three better methods are available to calculate statistics for data with non-detects.. *Maximum Likelihood Estimation* (MLE) solves a "likelihood equation" to find the values for mean and standard deviation that are most likely to have produced both non-detect and detected data. It requires the user to choose a specific shape for the data distribution, such as lognormal. Detected observations and the proportion of data falling below the MDL are used to fit the curve. But this method does not work well for data sets with less than 50 detected values, where one or two outliers may throw off the estimation, or where there is insufficient evidence to know whether the assumed distribution fits the data well. *Imputation* methods fill in values for censored data without assigning them the same value. As with MLE, the distribution must be specified, although imputation methods depend less on distributional shape. As a result, imputation methods generally perform better than MLE with small sample sizes or when the data do not exactly fit the assumed distribution. A non-parametric approach is the *Kaplan-Meier*. It does not require specification of distribution. It estimates the percentiles, or a cumulative distribution function for the data set. This method has been used primarily for "greater thans", so for low-level chemical concentrations, the data values must be individually subtracted from a large constant (flipped), before the software is run. Flipping data are necessary only because of the way commercial software is now coded; it may become unnecessary with future software versions as Kaplan-Meier becomes more widely used. A caution with this method is that the mean will be biased high when the smallest value in the data set is a non-detect (Helsel 2005).

According to Jones and Clarke (2005), the simple substitution methods listed above work better than the more complex techniques when sample sizes are very small. With small sample sizes, substitution of the MDL when less than 40% of the data are censored, or one-half of the MDL when more than 40% of samples are non-detects are methods that work reasonably well. In general, when the censored data exceed 60 to 80%, any statistical analysis is likely to result in unacceptably high error rates. This also assumes that the MDL values have been satisfactorily verified. The assignment of numerical values to non-detect data should be approached cautiously, with careful consideration given to uncertainties with this activity.

McDonald and LeClair (2006) used one-half the detection limit when 15% or fewer values were non-detects. Where the number of non-detects was between 15% and 50%, a probability plot regression was used to predict values in the lower tail of an assumed lognormal distribution for each non-detect. Comparative statistics were not calculated when the number of non-detects was greater than 50% of the measurements.

The USEPA (2000a) suggests that if the percentage of data below the detection limit is relatively low, then reasonably good estimates of means, variances and upper percentiles can be obtained. If simple substitution of values below the detection limit is proposed when more than 15% of the values are reported as not detected, consider using nonparametric methods or a test of proportions to analyze the data. But if the rate of censoring is above 50%, then little can be done statistically except to focus on some upper quantile of the variable distribution or on some proportion of measurements above a critical level above the detection limit. Table 6 provides a guideline for parameters to be used under different coefficients of variation and censoring levels.

Assumed Coefficient of Variation (CV)	Proportion of Data Below the Detection Limit	
	Low (<30%)	High (>30%)
Large: CV > 1.5	Mean or Upper Percentile	Upper Percentile
Medium: 0.5 < CV < 1.5	Mean or Upper Percentile	Upper Percentile
Small: CV < 0.5	Mean or Median	Median

Table 6. Guidelines for Recommended Parameters for Different Coefficient of Variations and Censoring (*USEPA 2000a*)

5. DISCUSSION

The technology for obtaining good data in aquatic monitoring programs continues to evolve – not only for sampling procedures, but also for a variety of other facets, including how to design programs to reduce costs while obtaining necessary data and how to assess and interpret results. Although this document attempts to provide guidelines for all of these, the field of quality assurance is not static, nor is this document. It should be updated periodically as new information becomes available.

To maintain consistency in ambient water and sediment sampling programs, project managers should understand the general ideas in this guideline and follow them as much as possible. However, professional judgment plays a major role in quality assurance planning, as it does in designing and executing a sampling program.

The Alberta measurement quality objectives listed in this document are based on actual QC data from WDS. They can be used as a reference to improve data quality, as it has not been optimum from some labs and studies in the past. But they should not be considered as final, and should be updated periodically. As well, each project manager should decide whether these are acceptable for his/her own project, or should be relaxed or strengthened. Project managers should use MQOs as guidelines – not so much to accept or discard data, but as flags for potential sampling or analytical problems to be investigated as quickly as possible. Therefore, QC data should be compared immediately with MQOs on arrival in the project manager's office, not left until the whole data set is available.

Although most labs used by AENV are accredited, the need arises occasionally to use labs that are not accredited or variables that are not yet accredited. This does not mean that the data are not as "good" as those from accredited labs, just that accreditation standards have not yet been applied to these. As well, accreditation does not guarantee that a lab will produce valid data, and for both of these reasons, a good QA/QC program is essential.

One of the most important unresolved issues is analytical detection limits. The literature is awash with disparate terminology, methodology and even how detection limits are reported. In particular, the laboratories used by AENV seem to have different ways of reporting non-detects. This could cause problems with long-term data sets if different labs are used over the course of a study. Masses of literature are available on how to statistically analyze and interpret a dataset with non-detects, and not all agree on how it should be done. This guideline document cannot resolve these issues. Professional judgment should be used to deal with these problems, or project managers should go to literature sources to find the best approach for their situation.

6. CONCLUSIONS AND RECOMMENDATIONS

Conclusions

- Quality assurance is the overall management of a sampling program to make sure that data are valid and usable. It includes field and laboratory standard operating procedures, as well as planning, documentation and validation of data.

- A good sampling design is an essential component of quality assurance.

- Quality control is a subset of quality assurance. It is used to measure the performance of the entire sampling program. Quality control measurements require a pre-specified quality standard.

- A key quality assurance question is how much error can be tolerated for the purpose of the study. This will vary with the study.

- Sediments are inherently more variable than water and require composite sampling.

- Assessment types of studies are more complicated to design than monitoring studies, and often require a testable hypothesis and the setting of data quality objectives.

- Alberta analytical laboratories use a variety of terms and methods for detection limits. Thus, it cannot be assumed that values near detection limit from one lab would be the same as those from another lab.

- Good communication among all parties involved in a sampling program and sample analysis is essential to reduce problems or resolve them if they occur. .

Recommendations

- Quality assurance and quality control should be included in all sampling programs done by and for the Alberta government.

- Project managers should complete a detailed sampling design for all ambient water and sediment sampling projects before they begin.

- The sampling design should include a statement of the problem, identification of information needs, a definition of study boundaries, a testable hypothesis (for assessment studies), MQOs and DQOs and the necessary elements to complete the study, including details of quality control samples.

- For most sampling programs, especially those to address environmental issues, splits, spikes, field blanks and occasional trip blanks should be done. Spikes and spike blanks (for pesticides) could serve the needs of several projects using the same lab.

- If possible, a QA coordinator, with statistical experience, should be available to help project managers with sampling design and statistical issues.

- Alberta Environment should conduct a laboratory performance test involving all accredited labs each year.

- A minimum of 10% of an analytical budget for a program should be allocated to QC samples. For certain studies (possible impacts to aquatic life or human health), as much as 35% of the budget could be allocated to QC.

- Variables known or suspected to be below analytical detection limits should not be analyzed in split/replicate samples.

- Split QC samples should not be sent to two separate labs, as it would not be possible to tell which results are right.

- Project managers should review quality control data as soon as it is available during a sampling program, so that quality problems can be investigated.

- For documented QC problems from a particular laboratory, the lab should review and reanalyze the samples at their own cost. All problems should be discussed with the lab.

- Alberta Environment should work with all the laboratories it uses to reach consensus on how detection limits are reported. Each laboratory should submit a detailed account of how their detection limit is done, including what it means statistically.

- For below-detection data, laboratories should report the actual value as well as the detection limit value, with appropriate uncertainties. It should be remembered that reported values below detection limit are highly uncertain, and generally should not be used in statistical tests.

- Field sampling techniques should be audited periodically to make sure the technologists are following standard operating procedures.

- Substitution methods for less than detection values should only be used when the number of samples is low or when the number of non-detects is low (<15%). When the number of non-detects is between 15-50% either use non-parametric methods or detailed techniques provided in the literature for statistical tests. If the

rate of censoring is above 50%, little can be done statistically except to focus on some upper quantile of the data distribution.

LITERATURE CITED

Alberta Environment. 2004. Alberta Environment laboratory data quality assurance policy procedures and guidelines. Environmental Assurance Div., Edmonton. 9 p.

Alberta Environment. 2006. Aquatic ecosystem field sampling protocols. 138 pages.

American Public Health Association, American Water Works Association, Water Environment Federation. 2005. Standard methods for the examination of water and wastewater, 21st Ed. Washington DC, USA.

Anderson, A-M. 2005. Overview of pesticide data in Alberta surface waters since 1995. Env. Monitoring and Eval. Br., Alberta Environment, Edmonton. 144 p + appen.

Anderson, A-M., T. Charette, B. Jackson, W. Koning, C. Teichreb, D.O. Trew, E. Watkins. 2005. Vision for Aquatic Ecosystem Monitoring.

CAEAL website: http://www.caeal.ca/ia_ilac.html.

Cavanagh, N., R.N.Nordin, L.W.Pommen and L.G. Swain. 1997. Guidelines for designing and implementing a water quality monitoring program in British Columbia. Resources Inventory Committee, British Columbia Ministry of Environment, Lands and Parks. http://ilmbwww.gov.bc.ca/risc/pubs/aquatic/design/index.htm

Clark, M. and P, Whitfield. 1994. Conflicting perspectives about detection limits and about the censoring of environmental data. Water Res. Bull. 30:6 p. 1063-1079.

Environment Canada. 2005. Pulp and paper technical guidance for aquatic environmental effects monitoring, Ch. 5.Measurement of supporting environmental variables. http://www.ec.gc.ca/eem/English/PulpPaper/Guidance/default.cfm

Eurachem. 2005. Quantifying uncertainty in chemical measurements. http://www.measurementuncertainty.org/mu/guide/index.html

Fisk, A.T., K. Hobbs, D.Muir. 2003. Interlaboratory quality assurance for the northern contaminants program. In *Contaminant Levels and Trends in the Biological Environment. Canadian Arctic Contaminant Assessment Report II.* Indian and Northern Affairs Canada: 111-128.

Government of British Columbia. 1998. Guidelines for interpreting water quality data. Ministry of Env., Lands and Parks. Prep. for Resources Inventory Committee. http://ilmbwww.gov.bc.ca/risc/pubs/aquatic/interp/index.htm

Hallock, D. and W. Ehinger. 2003. Quality assurance monitoring plan: stream ambient water quality monitoring. Washington State Dept. of Ecology Pub. No. 03-03-200. 28 p.

Helsel, D. 2005. More than obvious: Better methods for interpreting nondetect data. Env. Sci. and Tech. 39:20, p.419A-423A.

Jones, R. P., and Clarke, J. U. (2005). "Analytical chemistry detection limits and the evaluation of dredged sediment," ERDC/TN EEDP-04-36, U.S. Army Engineer Research and Development Center, Vicksburg, MS.

Katznelson, R. 1998. Tailoring of data quality objectives to specific monitoring questions. http://www.nwqmc.org/98proceedings/Papers/31-KAT.html

Lombard, S. and C. Kirchmer. 2004. Guidelines for preparing quality assurance project plans for environmental studies. Washington State Dept. of Ecology, Pub. No. 04-03-030, 47 p. + app.

McDonald, D. and D. LeClair. 2004. Methods and quality assurance investigations for trace metals data from the Long-Term River Network, 2003. Env. Monit. And Eval. Br., Alberta Environment. 77 p.

McEachern, P. and L. Noton. 2002. Overview of water quality in the Muskeg River Basin, July 1971 to March 2001. Science and Standards Br. Alberta Environment. W0201.

McQuaker, N. 1999. Technical evaluation on sample collection, handling, analysis and interpretation for trace level contamination in water. Aq. Effects Technology Evaluation Program Project 3.1.1, Natural Resources Canada, Ottawa.

Mueller, D., J. Martin, T. Lopes. 1997. Quality-control design for surface-water sampling in the National Water-Quality Assessment Program. Open-file report 97-223, U.S. Geol. Survey, Denver, Colorado.

Ontario Ministry of Environment. 1999. Protocol for the sampling and analysis of industrial/municipal wastewater. Municipal/Industrial Strategy for Abatement.

Statistics Canada. 2005. Statistics power from data. http://www.statcan.ca/english/edu/power/ch3/quality/quality.htm

USEPA. 1996. The volunteer monitor's guide to Quality Assurance Project Plans. EPA 841-B-96-003. http://www.epa.gov/owow/monitoring/volunteer/qappcovr.htm

USEPA. 2000a. Guidance for data quality assessment. Practical methods for data analysis. EPA QA/G-9. EPA/600/R-96/084. Office of Environmental Information, Washington DC.

USEPA. 2000b. Guidance for the Data Quality Objectives process. EPA QA/G4. EPA 600/R-096/055. Office of Environmental Information, Washington DC.

USEPA. 2001. EPA Requirements for quality assurance project plans. EPA QA/R-5. EPA/240/B-01/003. Office of Environmental Information, Washington DC.

USEPA. 2002a. Guidance for Quality Assurance Project Plans. EPA QA/G-5. EPA/240/R-02/009. Office of Environmental Information, Washington DC.

USEPA. 2002b. Guidance on Environmental Data Validation and Data Verification. EPA QA/G-8. Office of Environmental Information, Washington, D.C.

USEPA. 2002c. Guidance on choosing a sampling design for environmental data collection for use in developing a quality assurance project plan. EPA/QA/G-5S. Office of Environmental Information, Washington, D.C.

Wisconsin Dept. of Natural Resources. 1996. Analytical detection limit guidance and laboratory guide for determining method detection limits. Laboratory Certification Program. Publ-TS-056-96, Wisconsin.

Annotated Bibliography of Alberta Environment Reports Containing QA Information

Anderson, A-M. 1999. Water quality of the Battle River Technical Report. Part Three. Water Sciences Br., Alberta Environment. W9901.
> During a two-year water quality survey of the Battle River, five sets of duplicate samples and two sets of blanks were collected for quality control. Most variable concentrations were very similar in the duplicates, although a few had greater than 10% difference: BOD, silica, particulate carbon, dissolved and total aluminum and iron. The blanks had observable levels of a few constituents, although most were at or below the analytical detection limit.

Anderson, A-M. 2003. A survey of metals and trace organic compounds in sediments from Wabamun Lake and other Alberta lakes. Env. Monit. And Eval. Br., Alberta Environment, Edmonton.
> Metals were studied in Lake Wabamun sediments. The author found that the coefficient of variation (CV) for all the split samples was below 25% except for chloride. But the CV for individual sets of splits was above 25% for several variables. The variability among samples from the lake was much larger than that for the split samples.

Anderson, A-M. 2005. Overview of pesticide data in Alberta surface waters since 1995. Env. Monitoring and Eval. Br., Alberta Environment, Edmonton. 144 p + appen.
> Anderson (2005) conducted an overview of pesticide data for 1995-2002. She found that the quality of the pesticide database was similar to or better than that of other databases with respect to sample contamination, false positives, precision, and accuracy. Recovery rates were generally less than 100% and the likelihood of

false negatives was greater than false positives. As a result, reported concentrations and detection frequencies were likely biased low.

Casey, R. and P. Siwik. 2000. Concentrations of selenium in surface water, sediment and fish from the McLeod, Pembina and Smoky Rivers: Results of survey from fall 1998 to fall 1999. Interim report. Water Sciences Br. Alberta Environment. W0001.

> In a study of selenium in three northern Alberta rivers, the authors found no selenium in field blanks, and precision with coefficients of variation of 2 –22%, which was considered satisfactory. Spiked samples were analyzed to estimate accuracy. These had nearly 100% recovery in water and other media.

Casey, R. 2003. Disinfection by-products in the Wabamun Lake water treatment plant – treated water and Wabamun Lake, July 2002 to March 2003. Science and Standards Br., Alberta Environment.

> In a study of disinfection byproducts in treated and lake water from Lake Wabamun, different concentrations of chloroform were found in split samples sent to two laboratories. This suggests that concentrations from the regular sampling program for lake and treated water could be underestimated, although it would not change the conclusions of the study. As well, aldehydes were present in blank samples, suggesting contamination.

Cooke, S., P. Mitchell, L. Roy, L. Gammie, M. Olson, C. Shepel, T. Heitman, M. Hiltz, B. Jackson, S. Stanley, D. Chanasyk. 2002. Relationship between beef production and waterborne parasites (*Cryptosporidium* spp. and *Giardia* spp.) in the North Saskatchewan River basin, Alberta, Canada. CABIDF Project No. 97AB080, Prep. for Canada-Alberta Beef Industry Dev. Fund. AAFRD, Edmonton.

> Quality assessment and quality control were integral parts of the water quality component of the study of protozoan parasites in the North Saskatchewan River watershed. For most chemical variables and fecal coliform bacteria, precision was excellent, and blanks were usually clean. Splits sent to different laboratories were generally similar.

Hazewinkel, R. and L. Noton. 2004. Investigations of polychlorinated biphenyls in bottom sediments of the Bear-Wapiti-Smoky-Peace and Upper Athabasca River systems, 1989-2000. Env. Monitoring and Eval. Br. Alberta Environment. W0401

> PCBs were studied in sediments from several rivers in northern Alberta. Quality control samples included duplicate splits prepared in the field or lab, and submitted blind. Three sets of samples were sent to another lab to assess quality. Analytical precision among the split samples was generally good. There was poor agreement between data from the two labs, however. Possible reasons suggested for this were differing analytical detection limits or background contamination.

McEachern, P. and L. Noton. 2002. Overview of water quality in the Muskeg River Basin, July 1971 to March 2001. Science and Standards Br. Alberta Environment. W0201.

In an overview study of the Muskeg River, continuous recordings of dissolved oxygen, pH and conductivity were checked against laboratory or field values. The recorded data were either corrected or discarded, depending on the results. Field blanks were found to have levels of some metals that were sometimes higher than those in the river.

Mitchell, P. 1994. Effects of storm and combined sewer discharges in the city of Edmonton on water quality in the North Saskatchewan River. Surface Water Assessment Br., Alberta Environment, Edmonton.

One of the first extensive QC programs for an AENV study was that for an assessment of the impact of Edmonton storm sewer drainage on the North Saskatchewan River. This program found that many data had to be discarded because accuracy and precision were inadequate

McDonald, D. and D. LeClair. 2004. Methods and quality assurance investigations for trace metals data from the Long-Term River Network, 2003. Env. Monit. And Eval. Br., Alberta Environment. 77 p.

Quality control data for trace metals were evaluated from the Alberta Long-term River Network (LTRN) sampling program. The authors summarized split and spiked samples and evaluated data from two analytical laboratories. The study found that a large proportion of trace metal concentrations differed by more than 20% between the two labs, and that a number of blank samples contained metals at levels above method detection limits.

Quarshie, E. 1987. Interlaboratory quality control for the Water Quality Control Branch. Alberta Environment, unpub. MS.

A performance test for inorganic variables was conducted among labs used by AENV. A sample was collected from a river and split into a variety of aliquots, including several diluted with distilled water. Precision was calculated for each of the variables, and data from the labs were compared. Precision was generally good for most labs. One laboratory was found to have consistently high or low concentrations, and poor precision.

Appendix A. Quality Assurance in Alberta

Although standardized sampling procedures have always been used, the utilization of specific quality assurance/quality control procedures became common practice for Alberta Environment in the 1990s. It was aided by development of the Water Data System (WDS), a data storage system capable of identifying quality control samples from other samples. Alberta Environment's WDS was implemented in 1996, and since then, Alberta's water sampling programs regularly include quality control procedures. Over the past two years quality control has also been implemented in volunteer and provincial park lake sampling programs.

A new policy for laboratory data quality assurance in Alberta went into effect in 2001 (Environmental Assurance Division 2004). It states that all facilities supplying data required by an Alberta Environment Approval or Code of Practice must meet accreditation standards. This is intended to ensure that decisions regarding environmental management and sustainable development are based on best available data. The primary provider of accreditation is the Canadian Association of Environmental Analytical Laboratories (CAEAL). CAEAL is a full member of the International Laboratories Accreditation Standard Cooperation (ILAC), and is governed by accreditation standard ISO/IEC 17011. As of 2005, CAEAL is no longer associated with the Standards Council of Canada. As a general policy, surface water and sediment sampling programs conducted by AENV also use accredited laboratories, although new and emerging substances may not be accredited yet. You should remember, however, that laboratory accreditation does not guarantee good data. Many other aspects of QA can influence data quality.

One of the performance measures for the Government of Alberta accountability report, *Measuring Up*, includes results for the annual water quality index for the major rivers (Long-Term River Network sites).   The River Water Quality Index has been an official performance measure for several years. This province-wide program requires excellent data, which are scrutinized by staff of AENV and the Auditor General's office. Only accredited laboratories are used for acquiring data for the water quality index, and the AENV Surface Water Quality Monitoring Committee approves any changes to the sampling program. Regional limnologists review and verify all lab data from their region. As a result of the strict requirements of this performance measure, the data validation process for all sampling programs was redesigned and improved in 2002.

Most recent AENV reports contain QC results, with varying amounts of interpretation (*see Annotated Bibliography*). A few specific documents on quality assurance of particular sampling programs have been completed recently:

- McDonald and LeClair (2006) evaluated quality control data for trace metals from the Alberta Long-term River Network (LTRN) sampling program. They summarized split and spiked samples and evaluated data from two analytical laboratories. The study found that a large proportion of trace metal

---

concentrations differed by more than 20% between the two labs, and that a number of blank samples contained metals at levels above method detection limits.

- Anderson (2005) reviewed QA/QC measures for pesticides in Alberta surface waters for 1995-2002. Over this period, splits, blanks and spikes were done most years. Pesticide detection frequencies and concentrations were biased low, because recovery rates in spiked samples were usually less than 100%. This suggests that pesticide contamination in the environment is greater than that reported by the data. However, this is generally true for pesticide databases from other jurisdictions, and is not regarded as a weakness in the AENV database.

Quality Control Measures

Each regional limnologist or project manager presently designates quality assurance and quality control for surface water and sediment monitoring programs. Because there are no standard guidelines for study design, including numbers, types and locations of field QC samples, the QC program is inconsistent. In general, 5 to 10% of the analytical budget for any project is used for QC samples. For example, during pesticide sampling between 1995 and 2002, the number of QC samples (blanks, splits and spikes) ranged from 0 (1995) to 13% (1998) of the total number of samples collected, with an average of about 5% (Anderson 2005).

Some project managers specifically state where splits/replicates should be collected, while others leave it to the discretion of the sampling technologists. True splits are done more frequently than replicates. Splits are usually done as three samples (triplicate splits), although for some studies, the sample is split into two. Blanks represent one-third to one-half of the total QC samples, and these are field and trip blanks only.

Several quality control measures are applicable to all sampling programs. These include spikes, performance tests with reference materials and pesticide QC. Spikes and reference samples for nutrients, metals, chlorophyll *a* and pesticides are sent to various labs to assess performance, usually once per year. These are not done on individual studies, but as general quality assurance for a variety of programs. It is anticipated that the spiking program will be enhanced.

In the past, QC was not included on volunteer monitoring programs such as Alberta Lakewatch and Provincial Parks monitoring programs, but it will become a part of these programs in the future. An extensively trained field technologist trains Alberta Lakewatch samplers.

Quality control for field observations includes several calibration checks (see Alberta Environment 2006). McEachern and Noton (2002) provide an example of these procedures for datasonde installations. Continuous data for dissolved oxygen, pH and conductivity were checked against laboratory or field values determined from samples collected during installation and retrieval of the unit. If the paired values disagreed by a

constant number but the intervening data followed a consistent pattern, the datasonde values were corrected so that beginning and end values matched laboratory values. If a consistent pattern was not apparent, the datasonde values were discarded where they exceeded predetermined acceptance limits. Although Winkler dissolved oxygen determinations can be biased when high amounts of organic matter are present in the water, they are still used to calibrate datasondes as well as with membrane electrode instruments.

Detection Limits

The following are definitions of detection limits that are used in Alberta (from Ripp 1996, McQuaker 1999; Am.Public Health Assoc. 2005 and Jones and Clark 2005):

- *Minimum reported value* (MRV): at least 10 replicate analyses of a blank solution are carried out, and the standard deviation determined. The MRV is defined as 3 times the standard deviation. This is the lowest value a laboratory will report.

- *Method detection limit* (MDL) is based on data from samples containing the constituent of interest, which are processed through the complete analytical method. It is determined by adding a constituent to reagent water or matrix of interest to make a concentration from one to five times the estimated MDL. Seven to ten portions of this are analyzed and the standard deviation determined. A standard *t* distribution at 99% is used to calculate the MDL. MDLs are matrix, method and laboratory specific.

- *Reportable Detection Limit* (RDL) is the lowest value reported with confidence, The RDL is not a fixed multiple of the MDL, but the RDL must be greater than or equal to the MDL. RDLs would be considered once the results of the MDL were available (i.e. a calculated MDL of 4.82 might be rounded to 5.00 for the RDL).

- The *limit of quantitation* (LOQ) is the level above which quantitative results may be obtained with a specified degree of confidence. It is mathematically defined as 3.18 times the MDL. Limits of quantitation are matrix, method, and analyte specific.

- The *Practical Quantitation Limit* (PQL) is defined as "the lowest concentration that can be reliably achieved within specified limits of precision and accuracy during routine laboratory operating conditions. The PQL is generally 5 to 10 times the MDL, but it may be nominally chosen within these guidelines to simplify data reporting" (Jones and Clarke 2005). The PQL is significant because different laboratories will produce different MDLs even though using the same procedures, instruments and sample matrices. The PQL represents a practical and routinely achievable detection level with a relatively good certainty that any reported value is reliable (Am. Public Health Assn. 2005).

According to McQuaker (1999), an understanding of the meaning and significance of the different detection limits can aid in the interpretation of low-level data, such as very low concentrations of metal in water. For example:

- If a result < MDL then analyte not detected
- If a result is =/> MDL, but < RDL then analyte is present but result is not statistically significant
- If a result is =/> the RDL, but < LOQ then result is borderline statistically significant at the RDL.
- If a result is =/> the LOQ then result is statistically significant.

In the above case, RDL is the "reliable detection limit" which is two times the MDL. Each of the labs used by Alberta Environment has slightly different terminology and methodology for detection limit. Table A-1 lists major labs and their reporting criteria for detection limits. Further details are provided below.

Laboratory	Type of Analysis	Reporting Method	Comments
Alberta Research Council	Organics	actual value + <MDL	Values less than MDL are flagged
Alberta Research Council	Inorganics	>MRV, <MDL + SE for metals	Indicates analyte present but less than MDL
Maxxam	Inorganics	<RDL	
EnviroTest	Inorganics	< MDL	
University of Alberta	Inorg, Chl a	<MDL	LOQ estimated but not reported
Environment Canada	Inorganics (limited variables)	<MDL < PQL	j flag used for data between MDL and PQL

Table A - 1. Alberta analytical laboratories and their reporting criteria for detection limits in water. *See above for explanation of abbreviations.*

ARC Organics Group: "For the standard scans performed for AENV, we report the concentration (actual value) as calculated by the data processing software. The MDL value is listed beside the concentration on reports. Values less than the listed MDL can be included. A decision is made during the data review process to accept or reject a target compound hit. When we are confident the component is present, a concentration is calculated, and the data are reported with an appropriate flag when necessary; i.e. "X" indicates estimated value, concentration is less than the MDL. For data sent to the Water Data Server, the flag is switched to an "E" to match the required WDS code. We do not use floating or daily calculated MDLs. The MDLs are determined during method validation, and remain constant. This enables the use of VMV codes, and the ability to link the data to a particular method, and subsequently the MDL. We re-evaluate and update the MDLs whenever there is a significant change in the analytical methodology or instrumentation. If there are changes required, VMV and Method Codes would need to be updated. We set the MDLs (based on our experience) to a value we estimate will be a safe limit over a long period of time for the typical matrices being analyzed. We will normally give a safety margin of 3 to 5 times what we would expect to detect in most

samples for which the method is designed to be used. This safety margin allows for some fluctuations in extraction efficiencies, daily instrument sensitivity fluctuations, etc." (Grant Prill, ARC, pers. comm..)

ARC Inorganics Group: Only MDL and MRV are reported. The MDL is estimated from duplicate samples that are carried through the entire sample preparation procedure, and standard deviation of samples in the lower 20% of the sample range determined. The MDL is defined as 3 times this standard deviation. Typically upon development of a new method or initiation of new equipment, analysis of numerous blanks and samples (at least 30 of each) is carried out for preliminary evaluation. At least once in the first year of operation, all appropriate data are evaluated for a "final" assessment of MRV and MDL. For methods in use for several years, a regular review of collected data is carried out to see if any change is apparent. (F. Skinner, ARC, pers. comm.).

Maxxam: RDL is the lowest and the only detection limit reported. Data less than RDL is reported as <RDL or not detected. The RDL is not a fixed multiple of the MDL, but the RDL must be greater than or equal to the MDL. RDLs and MDLs are determined annually. (S. Raynard, Maxxam, pers. comm.)

EnviroTest: "We use the term detection limit (DL) on our test reports. In some cases the DL is the method detection limit (MDL). In other cases, the DL is greater than the MDL; i.e. the DL is set based on client needs. In all cases, the DL on our test reports is at least as high as the MDL. Our national laboratory accreditation body (CAEAL) considers the detection limit that appears on our analytical reports to be the RDL." We calculate MDLs at least every two years. (J. Weber, EnviroTest, pers. comm.)

University of Alberta: MDL is reported. They also use LOQ, but do not report this value right now. All of their detection limits are derived from method validations. Sensitivity changes are monitored during routine analyses, and if these should vary significantly, new detection limits would be calculated, and also after repairs or maintenance of instruments (M. Ma, U of A. pers. comm.).

Environment Canada Burlington: Results that are greater than or equal to the MDL and less than the PQL are reported. As well, if the internal standard has a recovery of <70%, the results will be flagged, and may be higher than what is reported. The detection limits are calculated once the method is determined. In every run a standard at the PQL level is run to ensure the instrument is working as it should. A continuing calibration check standard is run for every 10 samples. If the continuing calibration check standard is greater than the PQL, recovery must be +or- 20% of the actual value. Detection limits are determined again if any change is made to the method (S. Brown, Env.Can, pers. comm.).

Data Validation

Alberta Environment has set up a comprehensive methodology to validate data before they reach the electronic data storage system. The order of the following steps may vary

depending on the location of the regional office and the field technologist (*also see Appendix D*):

1. Field staff initiates the sample record in the database and confirms that sample header information is complete with valid codes and dates. QC samples are appropriately identified. Field sheets are sent to the project manager (photocopies, fax or whatever is effective and legible).

2. Project manager receives lab hard copies, then:

   - Ensures completeness based on project design,
   - Initiates re-runs and resolves any issues with field staff and/or contract labs,
   - Approves the invoice for payment,
   - Assembles the package under a validation cover sheet,
   - Informs contract labs of blind QC results, and
   - Returns the package to the field staff.

3. The field staff enters any outstanding data, measurement qualifiers and comments as per the project manager's instructions. Hard copies of the data are spot-checked against electronic data to make sure vmv codes and units are correct. The package is sent back to the project manager.

4. The project manager downloads data from WDS and checks any new data or corrections. The package is signed and sent to Data Management.

5. Data Management gives the package a final review for correctness and completeness of coding, vmv issues, etc. Sample records are migrated from the Temporary tables of the database to the Main tables. Data in Main are available to the public.

Appendix B. Alberta Environment Project Plan

The following is a list of the elements that each project should have identified and defined before sampling begins. These elements are for any surface water sampling program done by Alberta government staff, their contractors and partners. It has been modified from USEPA Quality Assurance Project Plan (QAPP) documentation. Some of the following QAPP elements have been combined to complement the way Alberta Environment currently plans and implements projects. It assumes that a professional scientist is designing the sampling programs.

Element 1 – Description of Project, Problem Definition/Background

- This should briefly describe the why, when, where statements and who will use the data.

Element 2 – List of Contacts/Responsibilities

- Project name and planned start and end dates.
- Organizations and main contacts involved in project plus their roles and responsibilities. This could include funding, program design, sample collection, samples analysis, data management, data evaluation and reporting
  - o AENV Project manager and techs
  - o Partner project manager, techs or contacts
  - o Analytical laboratories and contacts
  - o Data management contacts
  - o Alternates in all cases

  Note: Analytical labs should receive background on the project without including exact details.

Element 3 – Project Details

- The following information should be included:
  - o Manpower and budget estimates (details in Project Manager)
  - o How will sites be selected (GPS locations required)
  - o Runoff or storm event criteria
  - o Who will coordinate scheduling, what is plan if sampling is postponed, volunteers lose interest, etc.
  - o Details of stations, sampling frequency, variables required, analytical labs
  - o QC sample details
  - o Other sampling requirements: flow, etc.
  - o Logistics, special requirements (helicopter, boat, chain saws, etc)

Element 4 – Measurement Quality Objectives (MQOs)
- AENV will develop MQOs so that they are standard for all Surface Water projects for each matrix.
- Need MQOs for field variables, datasondes.
- What acceptance/rejection criteria are used in absence of MQOs?

Element 5 – Training Requirements / Certification

- AENV needs to develop training standards and where they will reside. What are the qualifications of a project manager, etc.
- Is there a need for 'provincial standards'?
- Determine if specialized equipment training or seasonal training is required.
- For volunteers, contractors, partners, what training will be provided by AENV? Identify trainers.
- Add link to any on-line training materials

Element 6 – Documentation and Records

- Field sheets, lab request forms, database coding information
- Equipment calibration/maintenance records
- Where will each record type be maintained? (not all are included in the data validation process)
- Links to on-line forms, all appropriate database codes (project number, station numbers, vmv codes, etc.)
- Water chemistry, datasonde, sediment, other types of sampling – each may have different requirements.
- Establish 'necessary requirements' for all documentation: organizations can use ours or develop their own as long as they contain the 'necessary requirements'

Element 7 – Sampling Methods Requirements

- Cite Sampling Methods and equipment SOP, cite Safety documentation
- Give links to on-line info: relevant sections of the SOP documents?
- Lists of lab bottles/preservatives/holding times
- GPS of "x" resolution required, 1:50,000 maps
- Establish 'Photo Documentation' standards. Store the photos on the server with the scanned version of field sheets.

Element 8 – Sample Handling

- Bottle labeling requirements
- Sample shipping requirements – packing, timing, confirmation of receipt at lab
- Clarification of filtering (field versus field lab versus contract lab)
- Disposal of sampling materials and hazardous goods (e.g. left over preservatives)

Element 9 – Analytical Methods Requirements

- Variable list mentioned under MQO's could be expanded to include recommended analytical methods and any special requirements (e.g. digestion of sediment metals samples)
- Methods need to 'meet or exceed' the AENV standard. For example, Dissolved Oxygen can be done using polarographic membrane, luminescent probe, or in some cases, modified Azide Winkler method. If a different method is to be used, it needs to be assessed before it can be considered to 'meet or exceed'.
- Identify seasonal considerations for methods (certain methods may work in open water, but not under ice).
- While this will often be pre-determined in most projects, what is the approach if the lab is not already selected? The project manager has to consider historic data, accreditation, availability of appropriate and proven analytical methods, cost, shipping time, turn-around time, electronic data transfer support, overall service, etc. This process would vary if project was a continuation or new.

Element 10 – Quality Control Requirements

- QC sample details will be outlined in the project details
- See Quality Assurance Guidelines document for QC sample details
- QC sample collection methods are in the Sampling SOP
- QC sample submission details (e.g. blind samples, third party spikes) are in the WDS Users Manual.
- Establish how QC results will be evaluated
- Determine the response/corrective action to contract labs.

Element 11 – Equipment Requirements

- List equipment that is required (must meet or exceed the specifications of AENV). Onus is on partner/contractor to prove their equipment meets these standards.
- Arrange equipment loan if required. Clearly state who is responsible for repair, replacement costs.
- Supply list of current suppliers of equipment, calibration standards, etc. (toolbox)
- Establish standards related to maintenance, calibration, 'changeout' of deployed units

Element 12 – Data Acquisition Requirements

- Identify any other types of data required for the project (i.e. not obtained through the monitoring activities), e.g. satellite photos, maps, land use info, WSC data, etc
- Discuss any limits on the use of these data from uncertainty about its quality.
- Identify who is responsible for gathering these data

Element 13 – Data Management and Validation

- Identify where all data components will finally reside (e.g. chemistry to WDS, flows and datasonde to TSM, scanned copies of spill data to a pre-arranged server location).
- Clearly identify data management process: from time of collection and lab analysis through validation to data storage and use.
- Establish who is doing what, i.e. who gets field sheets, who initiates sample record in database, etc.
- Describe how data are reviewed, accepted, rejected, qualified.
- Identify required data formats (contract lab, logged lake profiles, datasonde data, general chemistry format for data not entering WDS via lab loader)
- On-line link to WDS Users Manual

Element 14 – Inspection/Audit/Performance Assessment and Response Actions

- Should have a standard approach that would be fine tuned for each particular project
- AENV needs to evaluate field, lab, data management activities, organizations (contract labs, contractors) and individuals in the course of project. This could include site visits, refresher training sessions, audits of systems such as equipment and analytical procedures, audits of data quality (comparing results with MQOs).
- How will problems be corrected as identified through these assessments? (different level of response for a volunteer versus a contractor)
- Example: the project manager might decide that two side-by-side sampling trips will occur (one summer, one winter) after initial training has occurred, or that partner/contractor field equipment must be brought to the AENV test tank before sampling commences, etc.
- Contract terms should be enforced.

Element 15 – Project Status Reports

- Produce status reports – project status, results of internal assessments and audits, how QA problems have been solved. This should be formalized to deal with the contracted projects.
- The lines of communication should be improved by setting up protocol for feedback from partner or contractor for each field trip. How often do they report back and in what format? For example, they might be expected to email field sheets, site photos and a short description of how the trip went to the AENV project bio and tech.

Element 16 – Reconciliation with Objectives
- Did the data meet the planned objectives? Do any adjustments need to be made in the project? Do any data need to be discarded?

Element 17 – Reporting
- What reports are planned once data collection is complete?
- What type of report and covering how much time?
- Who will be receiving it?

Appendix C. Detection Limits

Detection limit is the smallest amount of a substance that can be detected above the noise in a procedure and within a stated level of confidence. Statistically, a data set with any observations recorded as being below a certain limit is called "left censored" or simply "censored." Terms for several detection levels are used in measuring environmental chemistry. In fact, Jones and Clarke (2005) cite eleven different terms for detection limits (some of which are interchangeable) and other publications use still different terms. Unfortunately, no standard definitions or procedures have been developed to make sure that detection limits are similar from one laboratory or one jurisdiction to another.

According to McQuaker (1999) the sample matrix can affect the analyte response and the detection limit. When detection limits are based on reagent water, the detection limit in different matrices can be altered by decreasing the analytical response and increasing variability in analyte recovery. MDLs may be calculated from blanks or matrix samples, depending on the laboratory.

An important criterion for selecting an analytical procedure is that it is done within a range of concentrations for which there is a simple linear relationship between the "true" and reported concentrations. Beyond this range, the relationship can become quite non-linear, and the random error grows rapidly larger as one moves out of the linear range. But with the growing importance of determining trace and ultra-trace amounts in the environment, more analyses are done at the lower end of the concentration range. If samples are analyzed too far into the zone of non-linearity, the data will be meaningless. The literature contains numerous procedures for determining analytical boundaries, called *detection limits* (Clark and Whitfield 1994).

The detection limit concept was developed to allow analysts to compare instruments and methodologies to help develop more powerful techniques. It was not intended to be used in data manipulation (Clark and Whitfield 1994). It is possible that a substance can be detected at concentrations well below the method detection limit. Statistically, any substance detected at a concentration equal to the MDL is 99% likely to be present at a concentration greater than zero. It also means that there is a 1% chance that a substance detected at the MDL will be considered present, when the true concentration is zero (Type I error). The MDL procedure is designed to protect against this type of error. But reporting data down to the MDL does nothing to control the possibility for false negatives (Type II error). False negatives are less of an issue for the environmental regulatory community, because nothing needs to be done if a substance is not detected. On the other hand, a false negative of a substance of concern could result in harm to aquatic life or other uses of the water. Another problem with the MDL is the assumption that precision is indicative of detectability. It is possible that a series of measurements may yield a high precision, but be biased either high or low. This would affect the calculated detection limit. These limitations are important to keep in mind when evaluating low-level data.

Data users should proceed with caution when interpreting data between the IDL and MDL (WI Nat Res 1996).

Analytical laboratories strive to provide their clients with reliable data. But this is not easy to achieve, especially for labs that have to process many samples each day while keeping costs and time required down. A majority of labs report data in a simple, understandable format – one result for each variable in the sample. Certified laboratories rigorously use standard methods that have been proven to be specific, precise, sensitive and unbiased, and they use standard protocols for the reporting of results. Analytical scientists view results below the limit of quantitation as somewhat dubious, and below the method detection limit very dubious. They suggest resources would be better spent developing more powerful methodologies rather than obtaining questionable results for thousands of samples (Clark and Whitfield 1994).

A fundamental principle of all analytical data is that they are estimates of true values. The information necessary to report the result of a measurement depends on its intended use, but it is preferable to err on the side of providing too much information rather than too little. Although true concentrations are never below zero, it is possible to report a dispersion of possible values that extend below zero. For example, when an unbiased measurement is made on a sample with no analyte present, one should see about half of the observations falling below zero, for example, observed concentration = -4.2+/-8 mg/L. This is a valid data report. There is no contradiction in reporting the best estimate and its associated uncertainty even when the result implies an impossible physical situation. It is essential when reporting a value for an analytical blank when it will be used to correct other results. The uncensored observation should always be available to the data user (Eurachem 2005).

Clark and Whitfield (1994) suggest that for electronic data, laboratories report results as pairs of numbers: the first value is the "official" result (i.e., censored), and the second the raw, unmodified result. Analytical labs should be encouraged to report QA/QC data as supplementary material. But environmental scientists should commit to distinguish clearly between analytical results and total assay results, both on paper and in computer files. This would allow statistical analysis of large data sets without resorting to infilling of non-detects with statistically fabricated data. Both laboratory and environmental scientists should provide explicit statements of analytical error.

Non-Detects
from NPSINFO Digest 1525 May 2001.

Adil N. Godrej, PhD
Associate Director
Occoquan Watershed Monitoring Laboratory
The Via Department of Civil and Environmental Engineering
Virginia Polytechnic Institute and State University
9408 Prince William Street

Manassas, VA 20110
U.S.A.

If you are evaluating the efficiency of a BMP based on comparison of influent and effluent EMCs, how do you treat values that are below detectable limits (BDL)?

This is truly a complicated (and much-debated) question, because the best answer is "It depends". Even the books on the subject do not discuss all the techniques that people have used and that have been defensible techniques (not necessarily the best) for a particular situation. With the availability of good mathematical software these days, using a host of techniques is possible. Rather than go into a detailed discussion of what's out there to handle this, I'll provide a few simple procedures that I use in approaching this problem. Of necessity, this is a simplified, and definitely not all-inclusive, discussion.

In most cases, we are talking about left-censored data, that is, some of the values are below the detection limit. (An example of right-censored data is coliforms readings that are above a certain limit, for example, >1600.) Assuming left-censored data, here are some of the questions I ask (not in an hierarchical order):

1. Is the detection limit (DL, although it may go by other names such as quantitation limit, etc.) above the modal value (equivalent to the median and mean for a normal distribution, basically this is the point where the 'hump' is in the distribution) of the distribution? Sometimes, people ask what the level of censoring is (that is, what percentage of all values is below the DL), but I think that, regardless of the level of censoring, most techniques that attempt to extrapolate the distribution from the above DL values will provide high-error results if the DL is above the modal value.

2. The second question I ask is, what is the level of censoring? If, say, 80% of the values are censored, then there are two possibilities: (a) if this is a long-term monitoring program then an effort should be made to lower the DL if we are dealing with a conventional pollutant, and (b) if this is data from a limited-study then it is time to hit the books and scratch your head and come up with what works best, realizing that the best may not be good enough. (More about (a) later in 5 (b).)

3.   To whom am I presenting the data? If to scientists, then I am free to use whatever fancy mathematical scheme I dream up, as long as I can defend it for that situation. If the data are to be presented to non-scientists or a mix of non-scientists and scientists, then I use the simplest techniques that will give me results close to what a fancier mathematical scheme might. Why? Because I find that if I use a fancy mathematical scheme on data presented to non-scientists, then I spend more time explaining and answering questions about the mathematics, and discussion of the data (which, presumably, is the main objective) often gets short shrift. For example, say I am computing loads for soluble phosphorus and my DL is low enough that I have about 10% censoring. At this point, a normal or a log-normal

distribution can be approximated with a straight line for that part below the DL without too much error, and I can use half the DL as a substitute for the censored values. The scientists in the audience will understand why I did it (I hope!), and the non-scientists will generally accept it without question because it sounds quite logical (and is logical within acceptable bounds of error for some cases).

4. What question is the data answering? If we are trying to get the worst-case scenario, then I often use a value equal to 99% of the detection limit for the censored data. This is blatantly wrong from a distribution perspective, so don't try this unless you have the right audience! Or, safer still, I provide a range of estimates with the low being censored data set to zero and the high being censored data set to the DL. This is of some use, but might get you accused of passing the buck if the data are highly censored, because then the range will be very large. Notice that I didn't say that I provide some sort of middle range value, because then you get back to what the distribution looks like below the DL and what value to use.

5. What constituent are you doing this for? This has two cases:
    a. Constituents such as TSS, nitrate, phosphorus, are both anthropogenic and naturally occurring, and it can be argued that even if the result is below the DL, there is probably some of it in the sample (that is, the sample belongs to a distribution). To get around this problem some folks are requiring labs to report all values, even those measured below the DL, and even those that are negative. The argument goes that getting a complete distribution (even though those values below the DL have tremendously high error-- which is why you have a DL--and may be meaningless from a nature perspective--less than zero concentration of TSS, for example, because the real value was so low that the weight of the filter after was less than its weight before the filtration step due to what the weighing balance can measure) gives a better mean value. Well, that might be true statistically speaking, but does beg the question as to why we are not using the median (instead of the mean) value for data that are in most cases log-normally distributed. If the answer to that is (i) very high censoring, then the real answer is to reduce the detection limit because there is so much error in the whole process with very high censoring that one might as well use half the DL instead of all values below the DL. If the answer is (ii) that we need a mean value because non-scientists understand averages, then I'd like you to stand in front of a group of non-scientists and explain a negative TSS/phosphorus/nitrate concentration ("You mean the water needs pollution to become clean and get to a zero concentration?", and, "Well, negative numbers may mean something to you scientists, but to me a negative is not a positive."). The public is generally suspicious of science that sounds shady, even though it may be more accurate, and we don't do ourselves any favors by coming across as "those scientists". Is the accuracy gained so much that it overshadows the lost scientific credibility in the public's eyes? Finally, in a general sense, asking lab people to report

all data, even that below the DL, is counter to the concept of having good QA/QC and well-defined DLs/PQLs/etc, and makes them very uncomfortable.

b. Constituents such as Atrazine are purely anthropogenic and do not occur in nature. Moreover, they are generally applied seasonally. So, if a value is below DL, the question is: Is it really present? When we report results from our lab, then, for such constituents, we use two codes for values below the DL. A code of <DL for those values where the compound was detected by the instrument but the concentration was below the DL; and a code of NF (not found) for those cases where the compound was not even detected. The person looking at the data then has the opportunity to ask him- or herself if the NF values should really be zeros, perhaps. For these non-anthropogenic, seasonally applied compounds, a fairly strong case can be made for a sort of bimodal distribution where a bunch of values lie at zero and there is some sort of more-regular distribution (log-normal, normal, etc.) above zero.

Granted that these are some selected thoughts, and there are many more facets to this issue. The main thing is to use the appropriate accuracy level and the appropriate techniques for the occasion--something you feel comfortable with and something you can defend for that application. That is sometimes easier said than done, and also is a function of experience. I'm sure that others in the NPS group will have lots to say on this issue.

Appendix D. Alberta Environment Data Validation Process

Data validation is a series of measures taken to ensure data quality. The data validation process for Alberta Environment surface water sampling programs was improved and updated in 2002. The following steps summarize this process.

All data in the "Temporary" tables of the database are considered 'preliminary' until this process is complete and they are migrated to the "Main" tables of the database.

Step 1 – Field Technologists

a. Confirm shipped samples arrived at lab.

b. Initiate sample in database within 2 weeks:
   - Create stations if necessary
   - Properly code all QC samples – add a comment describing QC sample (e.g. ARCV metals blank water poured off in field)
   - Enter sample comments
   - Load lake/reservoir profile files into Temp

c. Query the field data into a spreadsheet, confirm sample header info is correct. Hold original field sheets and mail/email/fax legible copies of field sheets to project manager.

d. Manually enter any other data that is not transferred electronically to WDS (and record on field sheet) as it arrives.

Step 2 - Project Manager

Based on field sheets and project details, the manager will initiate office validation:

   a. Gather lab data hard copies for a given time period into a common package, e.g. inorg, org, ultra trace Hg. Fill out a blue validation cover sheet for the package.

   b. Certify invoices and send to appropriate party for payment

   c. Query data using the Access Ad Hoc Tool. Confirm all samples on hard copy are in WDS.

   d. Confirm all samples and variables were collected as per project details.

   e. Check QC data against Measurement Quality Objectives. Check regular samples against normal concentrations for a particular station.

   f. Initiate reruns and/or inform lab of QC problems

   g. Resolve reruns when the rerun/QC file is returned from the lab – add comments to rerun/QC file and save in shared directory.

h. Labs will be instructed to produce re-issues of hard copies only if requested. (e.g., in the case of large errors). Minor changes/corrections can be written on the original copy **in red.**

i. Minor corrections, such as date/time, station – manager to note on hard copy **in red**

j. Major problems – sample swap, merged samples – resolve with appropriate lab, tech or data management staff.

k. Assemble package:

- List names/station numbers of sites or attach the 'query list' from "Sample Summary for Blue Packages" in the Ad Hoc Query Tool (enter a project number and date range to produce the appropriate list)

- Fill out the "Data" table and "Outstanding Data" area

- Initial and sign the blue validation cover sheet and attach printed copy of Rerun/QC file.

l. Send package to Field Tech. Include the field sheet copies in the package.

Step 3 - Field Tech
Complete field validation:

a. Query relevant data using the Ad Hoc Query Tool. Begin with a 'high level' review, and work down into the detail.

b. Review sample header info, stations, coding, etc. Make minor corrections identified by project manager on the hard copy. Re-query the data once these corrections are made.

c. Confirm all 'portions' for each sample are present in WDS (e.g. all the pesticides or metals) and that outstanding data are now entered.

d. Ensure WDS and hard copy are in agreement. Data must be reviewed in a spreadsheet for this check. Use Excel sorting techniques and macros, work 'on-screen'.

e. Reorganize the spreadsheet columns to match the groupings (e.g. field, metals). More macros are being developed to support this task.

f. Check all column headers, i.e. ensure that vmv, variable and unit match hard copy

g. Spot check one in 10 sample data to ensure all measurement values in WDS match the values on the hard copy

h. Watch for 'patterns' in the data – holes where there should be measurements, column with only one or 2 measurements, etc.

i. Note: Nearly all of the time the hard copy variable name, unit and method description are correct. If they don't match the database, there likely has

been a vmv translation problem, and the database is displaying incorrect data.

j. Contract labs LIMS are currently stable, so it is not necessary to check every measurement value.

k. Enter all re-run info (measurement qualifiers and comments) identified in the rerun/QC file, including sample comments for missing or deleted data. Note: if you are just fixing a vmv code, no meas qual or comment is required. Examples of rerun outcomes and resulting meas qualifiers and comments are found in the Rerun Template Directory.

l. Check for duplicate vmv/measurements (run Ad Hoc query "Find duplicate Meas in Temp")

m. Replace copies of field sheets in blue package with originals. Keep copies of field sheets for reference (scan and store on server)

n. Sign blue sheet and return to manager so they can review any data that were not entered earlier.

Step 4 – Project Manager

➢ Re-query data using Access Ad Hoc Tool to review extra data and give final approval to package. Sign and date blue package.

➢ Forward package to Data Management for final check and migration from Temp to Main.